



# UNIVERSIDAD DE LA RIOJA

## TRABAJO FIN DE ESTUDIOS

Título

Implementación de metodologías de dirección de proyectos en el desarrollo de tecnologías de la información para la captación de clientes en base a identificación de patrones de comportamiento

Autor/es

CRISTIAN OMAR CAMPUSANO NÚÑEZ

Director/es

FRANCISCO JAVIER MARTÍNEZ DE PISÓN ASCACÍBAR

Facultad

Escuela de Máster y Doctorado de la Universidad de La Rioja

Titulación

Máster Universitario en Dirección de Proyectos

Departamento

INGENIERÍA MECÁNICA

Curso académico

2017-18



***Implementación de metodologías de dirección de proyectos en el desarrollo de tecnologías de la información para la captación de clientes en base a identificación de patrones de comportamiento***, de CRISTIAN OMAR CAMPUSANO NÚÑEZ

(publicada por la Universidad de La Rioja) se difunde bajo una Licencia Creative Commons Reconocimiento-NoComercial-SinObraDerivada 3.0 Unported.

Permisos que vayan más allá de lo cubierto por esta licencia pueden solicitarse a los titulares del copyright.

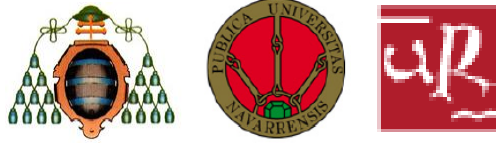
© El autor, 2018

© Universidad de La Rioja, 2018

[publicaciones.unirioja.es](http://publicaciones.unirioja.es)

E-mail: [publicaciones@unirioja.es](mailto:publicaciones@unirioja.es)

# MÁSTER INTERUNIVERSITARIO DE DIRECCIÓN DE PROYECTOS



## TRABAJO FIN DE MÁSTER

IMPLEMENTACIÓN DE METODOLOGÍAS DE DIRECCIÓN DE PROYECTOS  
EN EL DESARROLLO DE TECNOLOGÍAS DE LA INFORMACIÓN PARA LA  
CAPTACIÓN DE CLIENTES EN BASE A IDENTIFICACIÓN DE PATRONES DE  
COMPORTAMIENTO

AUTOR

CRISTIAN OMAR CAMPUSANO NÚÑEZ

DIRECTOR

DR. FRANCISCO JAVIER MARTÍNEZ DE PISÓN ASCACÍBAR

JULIO 2018

## Agradecimientos

A través de estas líneas expreso mi gratitud al Dr. Francisco Javier Martínez de Pisón Ascacíbar por su disposición de servir como asesor y mentor durante el desarrollo de este trabajo. Ha sido invaluable su aporte a mi carrera profesional.

Asimismo, agradezco profundamente al Dr. Eliseo P. Vergara González, Director Académico del Máster en la Universidad de La Rioja, por darme la bienvenida y permitir incorporarme al programa de clases de esta prestigiosa universidad.

A todos los profesores de la Universidad de Oviedo, Universidad Pública de Navarra y Universidad de La Rioja por compartir sus valiosos conocimientos.

Doy gracias a mi familia y amigos, quienes me han motivado a dar lo mejor de mí para la consecución de los objetivos que me he propuesto.

Agradezco inmensamente también al Ministerio de Educación Superior Ciencia y Tecnología MESCyT de la República Dominicana, bajo la dirección de la Lic. Alejandrina Germán. Su política de ofrecer facilidad de estudios a jóvenes, me ha permitido lograr una de las metas profesionales y culturales que más anhelaba.

# Resumen

Los proyectos de minería de datos para captar nuevos clientes necesitan de una dirección efectiva que garantice a las organizaciones el control, seguimiento y éxito en su gestión. Estos proyectos llevan asociados una serie de riesgos e incertidumbres en la determinación de los requisitos, que hace que su dirección requiera de metodologías especiales para lograr el mayor desempeño de los equipos de trabajo. En la actualidad, diversas investigaciones estudian la implementación del modelo de procesos más utilizado CRISP-DM a proyectos de minería de datos. No obstante, se confirma en los estudios empíricos la carencia de prácticas destinadas a procesos de gestión que contribuyan al logro de las metas de negocio. Este trabajo evalúa la aplicabilidad del marco de procesos Scrum como metodología ágil y los procesos que recomienda el Project Management Institute y sus guías de conocimiento. De igual modo, analiza el grado de efectividad que puede tener el uso de metodologías tradicionales predictivas para estos tipos de proyectos. A partir de los estudios realizados, se propone un marco de trabajo que ofrezca seguridad a las organizaciones al momento de abordar proyectos que contribuyan a aumentar su cartera de clientes.

**Palabras claves:** Minería de datos, CRISP-DM, Scrum, PMBOK, Metodologías ágiles

# Glosario de términos

<b>ANOVA</b>	En estadística, el análisis de la varianza (ANOVA, ANalysis Of VAriance, según terminología inglesa) es una colección de modelos estadísticos y sus procedimientos asociados, en el cual la varianza está particionada en ciertos componentes debidos a diferentes variables explicativas.
<b>Artificial Neural Networks</b>	Son sistemas de computación inspirados por las redes neuronales biológicas que constituyen los cerebros de los animales. Dichos sistemas "aprenden" a realizar tareas al considerar ejemplos, generalmente sin estar programados con ninguna regla específica de tareas.
<b>Association Rules Mining</b>	En minería de datos y aprendizaje automático, las reglas de asociación se utilizan para descubrir hechos que ocurren en común dentro de un determinado conjunto de datos.
<b>Big Data</b>	Es un concepto que hace referencia a conjuntos de datos tan grandes que aplicaciones informáticas tradicionales de procesamiento de datos no son suficientes para tratar con ellos y los procedimientos usados para encontrar patrones repetitivos dentro de esos datos
<b>Business Intelligence</b>	Hace referencia al uso de estrategias y herramientas que sirven para transformar información en conocimiento, con el objetivo de mejorar el proceso de toma de decisiones en una empresa.
<b>Classification trees</b>	Modelo de árbol de decisión donde la variable de destino puede tomar un conjunto finito de valores.
<b>Clustering</b>	Un algoritmo de agrupamiento (en inglés, clustering) es un procedimiento de agrupación de una serie de vectores de acuerdo con un criterio. Esos criterios son por lo general distancia o similitud. La cercanía se define en términos de una determinada función de distancia, como la euclídea, aunque existen otras más robustas o que permiten extenderla a variables discretas. La medida más utilizada para medir la similitud entre los casos es la matriz de correlación entre los $n \times n$ casos. Sin embargo, también existen muchos algoritmos que se basan en la maximización de una propiedad estadística llamada verosimilitud.
<b>CMMI</b>	Integración de modelos de madurez de capacidades o Capability Maturity Model Integration (CMMI) es un modelo para la mejora y evaluación de procesos para el desarrollo, mantenimiento y operación de sistemas de software.
<b>COCOMO</b>	El Modelo Constructivo de Costos (o COCOMO, por su acrónimo del inglés CONstructive COSt MOdel) es un modelo matemático de base empírica utilizado para estimación de costos de software.
<b>CRISP-DM</b>	Se trata de un modelo estándar abierto de procesos que describe los enfoques comunes que utilizan los expertos en minería de datos. Es el modelo analítico más usado.
<b>Data Science</b>	Es un campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no estructurados, lo cual es una continuación de algunos campos de análisis de datos como la estadística, la minería de datos, el aprendizaje automático y la analítica predictiva.
<b>Data Warehouse</b>	En el contexto de la informática, un almacén de datos (del inglés data warehouse) es una colección de datos orientada a un determinado ámbito (empresa, organización, etc.), integrado, no volátil y variable en el tiempo, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Se lo usa por reportajes y análisis de dato y se considera un componente meollo de la inteligencia empresarial.
<b>Datamarts</b>	Es una versión especial de almacén de datos (data warehouse). Son subconjuntos de datos con el propósito de ayudar a que un área específica dentro del negocio pueda tomar mejores decisiones.

<b>Deep Learning</b>	Es un conjunto de algoritmos de clase aprendizaje automático (en inglés, machine learning) que intenta modelar abstracciones de alto nivel en datos usando arquitecturas compuestas de transformaciones no lineales múltiples.
<b>DMCoMo</b>	Modelo Constructivo de Costos para proyectos de Minería de Datos.
<b>Electronic Word of Mouth (eWOM)</b>	Es una forma de marketing y puede volverse viral si el mensaje es persuasivo o lo suficientemente gracioso. En eWOM nos enfocamos en contactos de persona a persona que suceden en internet.
<b>FODA</b>	El análisis DAFO, también conocido como análisis FODA o DOFA, es una herramienta de estudio de la situación de una empresa, institución, proyecto o persona, analizando sus características internas (Debilidades y Fortalezas) y su situación externa (Amenazas y Oportunidades) en una matriz cuadrada.
<b>Gradient Boosting Machines</b>	Es una técnica de aprendizaje automático para problemas de regresión y clasificación, que produce un modelo de predicción en forma de un conjunto de modelos de predicción débiles, típicamente árboles de decisión.
<b>HTML</b>	Siglas en inglés de HyperText Markup Language (lenguaje de marcas de hipertexto), hace referencia al lenguaje de marcado para la elaboración de páginas web. Es un estándar que sirve de referencia del software que conecta con la elaboración de páginas web en sus diferentes versiones, define una estructura básica y un código (denominado código HTML) para la definición de contenido de una página web, como texto, imágenes, videos, juegos, entre otros.
<b>ISO 21500</b>	Es un estándar internacional desarrollado por la Organización Internacional de Normalización, o ISO a partir de 2007 y lanzado en 2012. Su objetivo era proporcionar una guía genérica, explicar los principios básicos y lo que constituye una buena práctica en la gestión de proyectos.
<b>Kanban</b>	Es un método para gestionar el trabajo intelectual, con énfasis en la entrega justo a tiempo, mientras no se sobrecarguen los miembros del equipo. En este enfoque, el proceso, desde la definición de una tarea hasta su entrega al cliente, se muestra para que los participantes lo vean y los miembros del equipo tomen el trabajo de una cola.
<b>Latent Semantic Analysis (LSA)</b>	Es una teoría y método para extraer y representar el significado de uso contextual de las palabras mediante cálculos estadísticos aplicados a un gran cuerpo de texto.
<b>Lean</b>	Es un modelo de gestión enfocado a la creación de flujo para poder entregar el máximo valor para los clientes, utilizando para ello los mínimos recursos necesarios, es decir, ajustados.
<b>Logistic regression models</b>	Es un modelo estadístico que generalmente se usa para aplicar a una variable dependiente binaria.
<b>Logs</b>	En informática, se usa el término log, historial de log o registro a la grabación secuencial en un archivo o en una base de datos de todos los acontecimientos (eventos o acciones) que afectan a un proceso particular (aplicación, actividad de una red informática, etc.). De esta forma constituye una evidencia del comportamiento del sistema.
<b>Machine Learning</b>	El aprendizaje automático o aprendizaje de máquinas es el subcampo de las ciencias de la computación y una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. De forma más concreta, se trata de crear programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos.
<b>Metodología ágil</b>	Es una metodología de gestión de proyectos que utiliza ciclos de desarrollo cortos llamados sprints para centrarse en la mejora continua del desarrollo de un producto o servicio, más que centrarse en la gestión del propio proyecto.

<b>Minería de datos</b>	Es el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos con el objetivo de encontrar patrones que nos puedan aportar información valiosa en la toma de futuras decisiones.
<b>Oracle</b>	Oracle Database es un sistema de gestión de base de datos de tipo objeto-relacional (ORDBMS, por el acrónimo en inglés de Object-Relational Data Base Management System), desarrollado por Oracle Corporation.
<b>PMBOK</b>	Guía de los fundamentos para la dirección de proyectos. Es un libro en el que se presentan estándares, pautas y normas para la gestión de proyectos. La última versión publicada es la 6ª, publicada el 6 de septiembre de 2017.
<b>Prince2</b>	Proviene del acrónimo en inglés PROjects IN Controlled Environments (PRINCE), es decir, convertir proyectos, que manejan una carga importante de variabilidad y de incertidumbre, en entornos controlados. Más que un conjunto de buenas prácticas, PRINCE2 propone una metodología de gestión de proyectos que cubre, mediante lo que se conoce como Temáticas, la Calidad, el Cambio, la estructura de roles del proyecto (Organización), los planes (Cuánto, Cómo, Cuando), el Riesgo y el Progreso del proyecto, justificado por un Business Case (o estudio de viabilidad) que debe ser revisado durante el ciclo de vida del proyecto y justificar en todo momento el proyecto como consecución de los beneficios esperados.
<b>Project Management Institute</b>	Es una organización estadounidense sin fines de lucro que asocia a profesionales relacionados con la Gestión de Proyectos. Sus principales objetivos son: Formular estándares profesionales en Gestión de Programas. Generar conocimiento a través de la investigación. Promover la Gestión de Proyectos como profesión a través de sus programas de certificación.
<b>Regression trees</b>	Modelo de árbol de decisión donde la variable de destino puede tomar valores continuos (por lo general números reales).
<b>Scrum</b>	Es un proceso de la Metodología Ágil que se usa para minimizar los riesgos durante la realización de un proyecto, pero de manera colaborativa.
<b>SEMMA</b>	Es un acrónimo que significa Muestra, Explorar, Modificar, Modelar y Evaluar (Sample, Explore, Modify, Model, and Assess). Es una lista de pasos secuenciales desarrollados por SAS Institute, uno de los mayores productores de software de estadísticas e inteligencia de negocios. Guía la implementación de aplicaciones de minería de datos.
<b>Sentiment Analysis</b>	(También conocido como minería de opinión) se refiere al uso de procesamiento de lenguaje natural, análisis de texto y lingüística computacional para identificar y extraer información subjetiva de los recursos.
<b>Social Learning Theory</b>	La teoría del aprendizaje social o TAS es la teoría de que las personas aprenden nuevas conductas a través del refuerzo o castigo, o a través del aprendizaje observacional de los factores sociales de su entorno. Si las personas ven consecuencias deseables y positivas en la conducta observada, es más probable que la imiten, tomen como modelo y adopten.
<b>Software</b>	Conjunto de programas y rutinas que permiten a la computadora realizar determinadas tareas.
<b>SPICE</b>	Modelo de evaluación (y mejora) de procesos software ISO 15504 SPICE es una adaptación para la evaluación de procesos de desarrollo software por niveles de madurez según la norma ISO/IEC 15504.
<b>SQL</b>	(por sus siglas en inglés Structured Query Language; en español lenguaje de consulta estructurada) es un lenguaje específico del dominio que da acceso a un sistema de gestión de bases de datos relacionales que permite especificar diversos tipos de operaciones en ellos.
<b>Stakeholders</b>	Hace referencia a una persona, organización o empresa que tiene interés en una empresa u organización dada.



<b>Support Vector Machines</b>	Son un conjunto de algoritmos de aprendizaje supervisado desarrollados por Vladimir Vapnik y su equipo en los laboratorios AT&T.
<b>Text Mining</b>	También conocido como minería de datos de texto, más o menos equivalente al análisis de texto, es el proceso de derivar información de alta calidad del texto.
<b>User-Generated Contents (UGC's)</b>	También conocido como contenido creado por el usuario (UCC), es cualquier forma de contenido creado por los usuarios de un sistema o servicio y disponible públicamente en ese sistema.
<b>Web Mining</b>	Es la aplicación de técnicas de minería de datos para descubrir los patrones de la Web.
<b>WebCrawlers</b>	Un rastreador web, indexador web, indizador web o araña web es un programa informático que inspecciona las páginas del World Wide Web de forma metódica y automatizada.
<b>XML</b>	Siglas en inglés de eXtensible Markup Language, traducido como "Lenguaje de Marcado Extensible" o "Lenguaje de Marcas Extensible", es un meta-lenguaje que permite definir lenguajes de marcas desarrollado por el World Wide Web Consortium (W3C) utilizado para almacenar datos en forma legible.

# Contenido

Agradecimientos.....	i
Resumen.....	ii
Glosario de términos .....	iii
Contenido .....	vii
Lista de Ilustraciones.....	xi
Lista de tablas .....	xii

## Capítulo 1

1. Introducción.....	2
1.1 Somos parte de un sistema de información.....	2
1.2 Problemática y objetivo general.....	2
1.3 Justificación del estudio a realizar .....	3
1.4 Hipótesis.....	4
1.5 Aporte y motivaciones personales .....	4
1.6 Estructura del documento.....	5

## Capítulo 2

2. Estado del arte.....	9
2.1 Evolución a Big Data .....	10
2.2 Técnicas de Minería de Datos para conocer más sobre los clientes.....	11
Intenciones de Compras.....	12
Clasificación de atributos .....	14
2.3 Procesos generales de Web Mining .....	14
2.4 Paradigma tradicional y reciente sobre el análisis de los datos.....	15
Predicción y Explicación .....	15

Acceso de Datos versus Independencia del científico .....	16
<b>2.5 Metodologías de gestión de proyectos utilizadas actualmente en Minería de Datos</b>	<b>16</b>
2.6 Evaluación de metodologías alternas para gestión de proyectos.....	18
2.7 Propuesta de trabajo .....	20

## Capítulo 3

<b>3. Metodología de la investigación.....</b>	<b>22</b>
--	-----------

## Capítulo 4

<b>4. Resultados y discusión .....</b>	<b>26</b>
4.1 Técnicas de Minería de Datos .....	27
4.2 CRISP-DM.....	30
Entendimiento del negocio .....	31
Entendimiento de los datos.....	32
Preparación de los datos.....	33
Creación de los modelos .....	33
Evaluación y pruebas.....	34
Despliegue o puesta en producción .....	34
Ventajas y Desventajas de CRISP-DM.....	35
4.3 Metodologías Predictivas.....	37
Requerimientos .....	37
Diseño .....	38
Implementación .....	38
Verificación .....	38
Mantenimiento.....	38
Evaluación de las metodologías predictivas.....	39
4.4 Recomendaciones de la guía del PMBOK con enfoque a proyectos de Minería de Datos.....	40
Ciclo de vida del proyecto.....	40

Fase del proyecto.....	40
Punto de revisión de fase .....	41
Procesos de la dirección de proyectos.....	41
Grupo de procesos de la dirección de proyectos.....	41
Área de conocimiento de la dirección de proyectos.....	42
<b>4.5 Comparación entre el Modelo de Procesos CRISP-DM y la Guía del PMBOK</b>	<b>48</b>
<b>4.6 Metodologías ágiles .....</b>	<b>50</b>
La prioridad es satisfacer al cliente de manera temprana y entregar continuamente software de valor.....	51
Aceptación de cambios incluso en fases finales del desarrollo del producto .....	51
Entregar software en funcionamiento con frecuencia, desde un par de semanas o un par de meses, con preferencia a escalas de tiempo más cortas .....	51
El personal de Negocios y desarrolladores deben trabajar juntos todos los días del proyecto.....	52
Construir proyectos alrededor de personas motivadas. Brindarles el entorno y apoyo que necesitan y confiar en que hagan el trabajo bien.....	52
El método más eficiente y eficaz para transmitir información a un equipo de desarrollo y dentro de él es la conversación cara-a-cara.....	52
Software funcional es la principal medida de progreso .....	53
Los procesos ágiles promueven el desarrollo sostenible. Los patrocinadores, desarrolladores y usuarios deberían poder mantener un ritmo constante indefinidamente.....	53
Atención continua a la excelencia técnica y al buen diseño aumenta la agilidad .....	53
La simplicidad, el arte de maximizar la cantidad de trabajo no realizado. Es esencial.....	53
Las mejores arquitecturas, requisitos y diseños surgen de equipos auto-organizados .....	54
A intervalos regulares, el equipo reflexiona sobre cómo ser más efectivo, luego optimiza y ajusta.....	54
<b>4.7 Marco de trabajo ágil Scrum.....</b>	<b>55</b>
Usos de Scrum.....	56
Teoría de Scrum.....	57
El Equipo Scrum.....	57

Eventos Scrum.....	58
Artefactos Scrum.....	59
<b>4.8 Comparación entre el Modelo de Procesos CRISP-DM y la metodología ágil Scrum</b>	<b>60</b>
<b>4.9 Propuesta de marco de trabajo para proyectos de Minería de Datos orientados a captación de clientes .....</b>	<b>61</b>
Fase de Iniciación: Enfoque predictivo .....	65
Fase de Planificación: Enfoque adaptativo .....	73
Monitoreo y Control: Enfoque adaptativo.....	79
Fase Sprint: Enfoque adaptativo .....	90
Fase Incremento: Enfoque adaptativo .....	93

## Capítulo 5

<b>5. Conclusiones y líneas futuras .....</b>	<b>97</b>
5.1 Conclusiones.....	97
5.2 Líneas futuras.....	99
Fuentes consultadas .....	100
<b>Anexos.....</b>	<b>1</b>

Anexo A.....	Acta de Constitución del Proyecto
1	

# Lista de Ilustraciones

<b>Ilustración 1.</b> Técnicas de Minería de Datos para explicar el pasado (Sayad, s.f.) ....	27
<b>Ilustración 2.</b> Técnicas de Minería de Datos para predecir futuros comportamientos (Sayad, s.f.).....	28
<b>Ilustración 3.</b> Metodología CRISP-DM, (Chapman et al., 2000) .....	31
<b>Ilustración 4.</b> Ciclo de vida en Cascada.....	37
<b>Ilustración 5.</b> Marco de Trabajo Scrum.....	56
<b>Ilustración 6.</b> Modelo de Incertidumbre y complejidad inspirado por el Modelo de Complejidad Stacey, (Project Management Institute, 2017b).....	62
<b>Ilustración 7.</b> Propuesta de Marco de Trabajo .....	63
<b>Ilustración 8.</b> Propuesta de Marco de Trabajo -Iniciación- .....	65
<b>Ilustración 9.</b> Roles de Equipo Belbin .....	71
<b>Ilustración 10.</b> Indicador de tipo de Myers-Briggs MBTI.....	72
<b>Ilustración 11.</b> Propuesta de Marco de Trabajo -Planificación- .....	73
<b>Ilustración 12.</b> Tabla Kanban, (Project Management Institute, 2017b) .....	76
<b>Ilustración 13.</b> Propuesta de Marco de Trabajo -Monitoreo y Control-.....	79
<b>Ilustración 14.</b> Gráfico de características, (Project Management Institute, 2017b).....	84
<b>Ilustración 15.</b> Diagrama BurnDown, (Project Management Institute, 2017b) .....	85
<b>Ilustración 16.</b> Diagrama BurnUp, (Project Management Institute, 2017b).....	85
<b>Ilustración 17.</b> Propuesta de Marco de Trabajo -Sprint- .....	90
<b>Ilustración 18.</b> Propuesta de Marco de Trabajo -Incremento-.....	93

# Lista de tablas

Tabla 1. Procesos Gestión de la Integración del Proyecto.....	43
Tabla 2. Procesos Gestión del Alcance del Proyecto.....	43
Tabla 3. Procesos Gestión del Cronograma del Proyecto .....	44
Tabla 4. Procesos Gestión de los Costos del Proyecto .....	44
Tabla 5. Procesos Gestión de la Calidad del Proyecto .....	45
Tabla 6. Procesos Gestión de los Recursos del Proyecto .....	45
Tabla 7. Procesos Gestión de las Comunicaciones del Proyecto.....	46
Tabla 8. Procesos Gestión de los Riesgos del Proyecto.....	46
Tabla 9. Procesos Gestión de las Adquisiciones del Proyecto .....	47
Tabla 10. Procesos Gestión de los Interesados del Proyecto .....	47
Tabla 11. Comparación Procesos PMBOK y tareas del modelo de procesos CRISP-DM.....	49
Tabla 12. Comparación Actividades Scrum y tareas del modelo de procesos CRISP-DM.....	60
Tabla 13. Modelo de Trabajo Metodología Propuesta.....	64
Tabla 14. Matriz de Asignación de Responsabilidades.....	69
Tabla 15. Roles de Equipo Belbin.....	72
Tabla 16. Matriz de evaluación de riesgos.....	78
Tabla 17. Matriz de respuesta y monitoreo a riesgos.....	78

# Capítulo 1

## Introducción, hipótesis y objetivos

“El valor de una idea radica en el uso que se le da” Thomas Alva Edison



# 1. Introducción

## 1.1 Somos parte de un sistema de información

Hoy en día muchas de las actividades que realizamos generan abundante cantidad de información. El uso del vehículo, el control de electrodomésticos que poseemos en casa, las compras que efectuamos, los viajes que realizamos, el uso de dispositivos electrónicos para fines de comunicación, entre otras tareas cotidianas; dejan un rastro digital que insinúa el marco de un perfil dentro de una red social. Los datos que resultan de estas actividades pueden utilizarse para formular una imagen de cómo es cada individuo; donde vive, hacia donde se dirige, qué necesita, cuáles son sus preferencias, entre otras peculiaridades. Todo esto ocurre a la vista de nuestros ojos de manera imperceptible si no contamos con las herramientas apropiadas para interpretar el valor de la información.

Estamos ante una era en que “el Big Data” nos está permitiendo descubrir lo que antes no habíamos podido ver. La tecnología nos ha permitido almacenar y procesar grandes cantidades de datos. Visualizar y analizar esos datos nos permitirán comprender el funcionamiento de sistemas complejos y detectar patrones sociales que permitirán a las organizaciones tomar decisiones estratégicas que contribuyan a su crecimiento en el mercado.

Los seres humanos se han convertido en los nuevos nodos de la Internet, permitiendo así que seamos contribuyentes de información para la red digital a la que pertenecemos.

## 1.2 Problemática y objetivo general

En tiempos recientes, ha surgido en las organizaciones la necesidad de emprender proyectos que les permitan anticiparse ante las demandas del mercado y afrontar la extenuante competencia de otras compañías.

Las principales cuestiones que se plantean los directivos de empresas resultan ser: cuáles son las características primordiales que deben poseer sus productos o servicios, qué opiniones tienen sus clientes de la empresa, si están satisfechos u optarán sus clientes por preferir a la competencia. Con el beneficio que ofrecen las

nuevas tecnologías y herramientas de minería de datos, estas preguntas pueden ser respondidas.

La gestión de proyectos de minería de datos requiere del empleo de metodologías adecuadas a los procesos relativos con la identificación, análisis e interpretación de datos por equipos de trabajo de disciplinas diferentes y disímiles funciones. En la actualidad no existe una metodología o marco de trabajo estándar destinado específicamente para administrar todas las fases de estos tipos de proyectos. Las organizaciones, al utilizar metodologías tradicionales de desarrollo de software convencionales para gestionar proyectos relacionados a minería de datos, generalmente no consideran las particularidades asociadas a estos tipos de proyectos, lo cual repercute negativamente en la gestión efectiva de los equipos de trabajo y recursos disponibles.

Exceder en presupuesto al momento de afrontar los proyectos, fracaso en la ejecución de los procesos de minería de datos al no evaluar correctamente los riesgos, perder competitividad frente a otras empresas por no identificar efectivamente la demanda del mercado, son algunas de las consecuencias que conlleva no gestionar los proyectos bajo un marco de trabajo apropiado.

El enfoque de este estudio se basa en evaluar las diversas metodologías que pueden utilizarse en proyectos que impliquen tareas de minería de datos con el objetivo de que las organizaciones puedan aumentar su cartera de clientes.

### 1.3 Justificación del estudio a realizar

Para abordar proyectos de minería de datos puede considerarse las metodologías orientadas a ejecución de procesos en cascada que determinan modelos lineales de diseño de software bajo un esquema secuencial de etapas. Esta manera de trabajar resulta ser de las más tradicionales en gestión de proyectos, en la que se establece un firme enfoque a las actividades del plan de trabajo hasta cumplir con el objetivo claramente definido en etapas tempranas. No obstante, los proyectos de minería de datos se caracterizan generalmente por iniciar con poca definición de requisitos y cambios recurrentes durante el desarrollo del mismo, lo cual crea la necesidad a las organizaciones de optar por utilizar metodologías más flexibles.

Diversas metodologías existen en la actualidad para trabajar con proyectos de minería de datos, sin embargo, presentan debilidades al no ofrecer detalles para todos los procesos asociados.

Las organizaciones deben encaminarse a crear un entorno laboral propicio para la ejecución de estos proyectos y favorecer la cooperación en diversos equipos de personas tras un fin determinado. Deben poder monitorear la evolución del trabajo en cada etapa con el objetivo de controlar sobrecostos, cambios significativos en el alcance y aparición de nuevos riesgos antes de que sea demasiado tarde.

## 1.4 Hipótesis

1. Preparación de un marco de trabajo integral que incluya las flexibilidades que permiten tener las metodologías ágiles, el enfoque detallado de las metodologías diseñadas específicamente para proyectos de minería de datos y las buenas prácticas descritas en el PMBOK.

La consideración de procesos de monitoreo y control en las diversas etapas del proyecto para fines de preservar el costo estipulado y alcance dentro de límites asiduamente supervisados y aceptados son esenciales para que las organizaciones puedan gestionar de forma eficaz sus proyectos de minería de datos.

2. Como alternativa para la gestión de proyectos de minería de datos se puede considerar las mejores prácticas descritas por el PMI respecto a las metodologías ágiles. El cumplimiento de estas recomendaciones y el uso de una metodología ágil como lo es Scrum ha de ser suficiente para una dirección eficaz ya que estas metodologías han resultado ser exitosas en los desarrollos de proyectos de software de alta criticidad, con constantes cambios e involucramiento de los equipos que lo conforman.

## 1.5 Aporte y motivaciones personales

Este trabajo de investigación representa un aporte a la comunidad empresarial y científica que se fundamenta en las mejores prácticas de dirección de proyectos; específicamente a los relacionados a comprender lo que esconden los datos capturados en las webs comerciales, los cuales revelan patrones de conductas en los individuos de nuestra sociedad dentro de determinados mercados. Contribuye a desarrollar nuevas técnicas de dirección para beneficio colectivo.

El análisis de los estudios que publican los científicos en la literatura, la guía de las mejores prácticas descrita en el PMBOK, las recomendaciones descritas por el PMI para los proyectos ágiles, además de los materiales impartidos en clases han de servir como fundamento para las recomendaciones a ser expuestas en este trabajo.

En términos personales, el principal interés que me atañe en consecuencia al desarrollo de esta labor académica corresponde a enriquecer mis habilidades para la gestión efectiva de proyectos, profundizando en los conceptos discutidos en clases y contribuyendo de esta forma a un notable crecimiento en el ámbito profesional. Otro aspecto que espero cultivar es la redacción correcta de documentos técnicos con el objetivo de que sean entendibles en una comunidad de profesionales asociados a la materia. Algo más que espero perfeccionar es: mi capacidad de análisis de artículos científicos y evaluación de las posturas de sus autores, de forma tal que mis argumentos profesionales cuenten con fundamento científico.

Lo fascinante de la evolución en las tecnologías que se utilizan para estudiar aspectos psicológicos de las personas y descubrir en los resultados comportamientos inimaginables, resulta ser en mí una especial motivación por la cual dedicar tiempo y esfuerzo en este trabajo final. Por tal razón he querido conjugar en el tema de estudio que he elegido para este trabajo de fin de máster: la psicología social que tanto me apasiona, la tecnología que ha sido mi área de estudios por años y el enfoque de dirección de proyectos con el fin de crecer a nivel profesional.

## 1.6 Estructura del documento

Este trabajo de fin de máster está organizado en capítulos, para los cuales se presenta a continuación una descripción de su contenido.

- Capítulo 1: Este primer capítulo introduce al lector sobre la problemática que afecta a las organizaciones al momento de abordar proyectos que contribuyan a la captación de nuevos clientes utilizando tecnologías de Minería de Datos. Presenta el objetivo general por el cual se realiza el estudio. En adición, se presentan las hipótesis iniciales para la situación que viven las empresas al experimentar fracasos en sus proyectos. Se expresa también cual es el aporte y motivaciones personales sobre el autor para realizar este trabajo.
- Capítulo 2: Esta sección presenta el estado del arte sobre las metodologías conocidas sobre dirección de proyectos de Minería de Datos, y las técnicas utilizadas para conocer los patrones de comportamientos de las personas en el ámbito comercial. Se evalúa cuales son los paradigmas sobre el análisis de

los datos. A partir de los hallazgos, se plantea cual es la propuesta de trabajo a desarrollar.

- Capítulo 3: Se plantea la metodología empleada para realizar el análisis y documentación del trabajo.
- Capítulo 4: Este capítulo presenta una evaluación sobre las mejores prácticas sobre dirección de proyectos y su aplicabilidad en proyectos de Minería de Datos.
  - Se presenta una clasificación de las técnicas de Minería de Datos más conocidas, y delimita el enfoque del estudio a proyectos relacionados a técnicas de creación de modelos predictivos.
  - Se describen los conceptos del modelo de procesos CRISP-DM; modelo mayormente aceptado para desarrollo de proyectos de Minería de Datos, y se analiza cuáles son sus ventajas y desventajas.
  - Se evalúa la aplicabilidad de metodologías de dirección de proyectos con ciclo de vida predictivo.
  - Se muestra el enfoque de la guía del PMBOK del Project Management Institute frente al modelo de procesos CRISP-DM.
  - Se presenta una comparación entre las mejores prácticas de la guía del PMBOK y las etapas y actividades del modelo de procesos CRISP-DM.
  - Se evalúa el enfoque de las metodologías ágiles frente al modelo de procesos CRISP-DM.
  - Se describen los conceptos del modelo de procesos ágil Scrum.
  - Se presenta una comparación entre las fases y actividades de Scrum frente a las etapas y actividades del modelo de procesos CRISP-DM.
  - Se propone un marco de trabajo para la gestión de proyectos de Minería de Datos para captación de clientes.

- Capítulo 5: Se presentan las conclusiones y líneas futuras sobre trabajos de investigación relacionados al tema de estudio.

# Capítulo 2

## Estado del arte y propuesta de trabajo

“Sólo existe una fuente de conocimiento: la experiencia” Albert Einstein

## 2. Estado del arte

En el ámbito comercial, la publicidad conocida como *boca a boca* ha sido por largo tiempo un referente en las organizaciones a la hora de centrar sus estrategias de ventas y captación de clientes. Hoy en día la internet hace posible que la influencia de las personas no incida únicamente en un espacio demográfico o entorno social limitado, sino más bien, permite que la opinión que expresa una persona en una plataforma de comercio electrónico pueda afectar positiva o negativamente la reputación de la empresa comercial a nivel de toda su red de clientes. En el ambiente altamente competitivo en que se desempeñan las empresas, debe aprovecharse cada oportunidad de mejora para no cesar en sus funciones de negocio. Las organizaciones tienen acceso vía internet a inmensa cantidad de información que comparten sus clientes sobre los productos o servicios que ofrecen, sin embargo, este insumo de primera mano normalmente no es aprovechado de manera eficiente con miras a mejorar la propuesta que brindan a su mercado destino. Técnicas de mercadeo basadas en encuestas, entrevistas, grupos focales, entre otras; han ido complementándose en los últimos años con herramientas tecnológicas en pleno desarrollo para detectar patrones de comportamientos intrínsecos de los clientes que no suelen revelarse con métodos tradicionales. La aplicación de sistemas innovadores que consisten en procesos de minería de datos se han convertido en componentes fundamentales para las empresas posicionadas en la web con acceso a las impresiones que tienen sus clientes respecto a los productos o servicios que distribuyen.

El uso de procesos para descubrir información útil a partir de los datos que comparten los clientes en la web y que regularmente se presentan como no etiquetados o no clasificados, en vasta cantidad, heterogéneos (en formato de video, texto, audio, etc.), semiestructurados (*xml*, *html*, entre otros), cambiantes en el tiempo o con un alto número de dimensiones representa un reto a los científicos de datos por las complejidades inherentes asociadas a la psicología social y a las estructuras en que se almacenan los datos, (Bandyopadhyay y Pal, 2007). Razones como éstas, hacen que las organizaciones teman abordar proyectos de este tipo, aunque cada vez más vemos cómo evolucionan herramientas que hacen más fácil la explotación de los datos para ganar más conocimientos sobre la base de clientes a la cual se dirigen las empresas.



La información contenida en la Web puede categorizarse en general de la siguiente manera:

- Contenido Web: Que representa la información almacenada de texto, imágenes, sonido, registros estructurados como listas y tablas.
- Estructura Web: Componentes que definen las conexiones entre el contenido de la Web, por ejemplo: hipervínculos y etiquetas.
- Uso Web: Información que describe la interacción de los usuarios con la Web, por ejemplo: HTTP, logs, entre otros.

Como parte de este trabajo, el estudio está enfocado al aprovechamiento de la información y conocimiento que se puede extraer a partir de los Contenidos Web en formato de Texto almacenados en registros estructurados producto de las impresiones que generan los usuarios como parte de sus experiencias comerciales posterior a la adquisición o uso de productos y servicios; permitiendo que otros estudios se basen en los demás contenidos de los cuales se puede obtener informaciones valiosas para beneficio de las organizaciones.

## 2.1 Evolución a Big Data

El concepto y conjunto de estrategias que proveen las herramientas de Inteligencia de Negocios han aportado desde su mayor auge en los años 80's conocimiento a partir de las informaciones disponibles a los directores departamentales para toma de decisiones claves en las organizaciones. Las metodologías, aplicaciones y tecnologías utilizadas como parte de los procesos de Inteligencia de Negocios para recopilar, depurar y transformar datos de los sistemas transaccionales en información estructurada para fines de crear reportes y análisis a partir de ellos, ha contribuido a que las empresas conozcan más de la evolución de sus operaciones. A medida que el mercado es más competitivo y las empresas cuentan con mayor retroalimentación de sus clientes, procesos más precisos y eficaces surgen para poder estar a la vanguardia en el sector. Según Ishikiriya et al., (2015), los procesos de análisis de datos y el concepto de *Big Data* nacen a partir de la necesidad de sacar el mayor provecho de la información disponible, provocando que este sea el nuevo enfoque y no tanto las herramientas tradicionales de Inteligencia de Negocios. El análisis de los datos conduce a la organización a tomar decisiones más inteligentes, desempeñar operaciones más eficientes, obtener mayores ganancias y mantener clientes satisfechos.

## 2.2 Técnicas de Minería de Datos para conocer más sobre los clientes

Teniendo en cuenta que es mucho más costoso para una empresa adquirir nuevos clientes frente a mantener los existentes, la organización debe tener la astucia de concentrar sus esfuerzos en evitar que su base de clientes detecte sus debilidades antes de que sea demasiado tarde y no pueda seguir compitiendo en el mercado. Técnicas como *Deep Learning*, *Gradient Boosting Machines*, *Support Vector Machines* (SVM), como también las que presentan en su estudio Tiwari et al., (2010): *Classification and regression trees* (CART), *Logistic regression models* (LRM) y *Artificial Neural Networks* (NN), resultan ser en gran medida utilizadas para predicciones de deserciones de clientes en las organizaciones. Estas técnicas de modelos de aprendizajes supervisados, así como otros enfoques que se han presentado en la literatura ayudan bastante con la predicción de deserciones de clientes, sin embargo según Pinheiro y Helfert (2010), no consideran la influencia que pudieran tener unos clientes sobre otros en la web. Análisis de incidencia en redes sociales pueden utilizarse para optimizar el grado de conocimiento que las organizaciones tienen sobre sus clientes. Es posible que el cliente que figure de menos valor en términos de beneficios a la organización, represente ser un nodo en la red de clientes de alta influencia sobre los que les rodean; lo cual puede significar un alto impacto en la lealtad de los clientes. El uso combinado de técnicas tradicionales y nuevos enfoques para explotar *Big Data*, perfeccionan las fortalezas de cada una y compensan las debilidades de otra, (Mahmoodi et al., 2017).

Con el apoyo de técnicas de *Text Mining* para descubrir informaciones de alto nivel a partir de grandes volúmenes de datos de texto, las organizaciones pueden extraer más valor de lo que puede verse a simple vista. El término *Text Mining* se refiere al proceso de extraer información útil y significativa a partir de texto no estructurado (Netzer et al., 2012).

Las informaciones pueden ser presentadas de forma útil a potenciales usuarios, que podría ser un analista, un director, un gerente de proyecto, el presidente de la compañía u otra persona con la autoridad suficiente para tomar decisiones relevantes dentro del progreso de la compañía. Como afirman Boyd y Pennebaker (2017), es posible identificar como producto de procesos de *Text Mining*: aptitudes, emociones y patrones de comportamientos que tienen las personas. El uso de encuestas para capturar las disposiciones de los clientes, han resultado no ser del todo precisas, debido a que las personas suelen expresar un modelo de cómo creen que son y no cómo realmente se comportan en determinadas situaciones; a lo largo de los años psicólogos han confirmado este comportamiento, (Tausczik y Pennebaker, 2010). En

ese orden, con los endosos “*Likes*” en redes sociales se puede deducir perfiles en la personas de manera muy precisa, (Kosinski et al., 2013). Teniendo esto en consideración, no solo es recomendable el uso de técnicas tradicionales, modelos estadísticos para determinar patrones en los clientes o procesos de Inteligencia de Negocios para descubrir informaciones valiosas en los datos; sino también resulta beneficioso apoyarse en procesos de *Text Mining* para conocer específicamente qué opinan los clientes y además determinar sus aptitudes inherentes a lo que han expresado; utilizando por ejemplo técnicas de *Sentiment Analysis (SA)*. La meta de aprender a partir de la base existente de consumidores contribuirá a que la organización desarrolle mejores propuestas dirigidas a potenciales nuevos clientes.

*Opinion Mining* o como también se le conoce: *Sentiment Analysis (SA)*, es un conjunto de técnicas que se utilizan para el procesamiento de las opiniones expresadas en la web con miras a clasificar las actitudes de los usuarios principalmente en las páginas comerciales y así interpretar qué les gusta y qué no sobre sus ofertas mercadológicas. Estos sentimientos se pueden utilizar para medir el pulso de los clientes de cara a la propuesta comercial de las organizaciones. Según Kurada y Kanadam (2018), los procesos de SA resultan ser efectivos si son empleados correctamente, ya que se puede lograr a interpretar las polaridades positivas, negativas y neutras a partir de *sets* de datos utilizando el lenguaje de programación R, de igual modo utilizando algoritmos de *Machine Learning* Basados en Reglas, (Ahmed y Danti, 2016).

## Intenciones de Compras

Parte de las actitudes que pueden detectarse de los clientes a partir del estudio psicológico por medio de lo que han compartido en las plataformas comerciales de venta de artículos, foros y redes sociales, son las intenciones de compra. Tal como expresan Chen et al., (2017), basándose en la Teoría *Social Learning Theory* es posible detectar las intenciones de compra de los clientes. Las Social Commercial Components (SCCs) presentes en las plataformas de comercio electrónico inciden en la predisposición de compra por parte de los consumidores, tanto en dimensiones emocionales como en cognitivas. Los resultados de este estudio sugieren que las dimensiones emocionales y cognitivas de los consumidores son los principales predictores para la intención de compra.

## Apreciación emocional

La dimensión emocional consiste en apreciaciones basadas en los sentimientos, reacciones, motivaciones e impresiones que los individuos experimentan en relación con el objeto que le atrae. En el comercio social, la apreciación emocional mide las experiencias hedónicas como sentirse feliz, bueno, relajado, agradable, o satisfactorio con las interfaces web y la promoción del artículo o servicio.

## Apreciación cognitiva

La apreciación cognitiva es la dimensión asociada al aprendizaje, que puede describirse como el conocimiento o habilidades que desarrolla la persona para comprender el objeto de estudio. El esfuerzo de la persona consiste en construir un modelo pragmático y útil para uso cotidiano del artículo o servicio que evalúa.

El estudio que realiza Chen et al., (2017) revela que la apreciación cognitiva resulta ser de mayor influencia sobre la apreciación emocional al momento de evaluar la elección de un producto o servicio. La investigación se ha realizado en base a entrevistas a un segmento de clientes de la empresa Taobao en China (<https://daren.bbs.taobao.com>). Teniendo en consideración estos resultados, las organizaciones deben orientar sus tácticas a estos factores que inciden en el comercio social, principalmente a la dimensión cognitiva, la cual se apoya en la tendencia de los contenidos generados por los usuarios en las plataformas de comercio electrónico.

El contenido que generan los usuarios en las plataformas de tiendas en línea, como se les conoce: *Electronic Word of Mouth (eWOM)*, representa un valor significativo ya que denota influencia de compra por parte de otros usuarios y demanda de productos o servicios en la sociedad, (Xie et al., 2014). Cabe destacar que, incluso el diseño visual de la presentación de los comentarios de los usuarios y si son fáciles de leer tienen incidencia en la decisión de compra para quien estudia las opiniones de los demás. En adición a esto, el uso del tipo de letras en los comentarios que escriben los usuarios clientes por adquisición de productos vía tiendas en la web se considera en menor medida como factor de incidencia en la decisión de compra, (Huang et al., 2018).

## Clasificación de atributos

Más allá de identificar las emociones, patrones de comportamiento de los usuarios a través de los comentarios en los cuales expresan sus quejas o satisfacciones sobre un particular producto o servicio, es posible detectar específicamente qué atributos han sido los que han provocado la reacción de los clientes. Según Gaurav y Kumar (2017) resaltan en su estudio cuáles son las características más valoradas por los usuarios en las opiniones expresadas en una plataforma comercial, mientras que Xu et al., (2017) utilizan *Latent Semantic Analysis (LSA)* y regresión basada en las revisiones textuales de los usuarios para examinar las percepciones de los atributos o características específicas que los motivaron a expresarse en respuesta a su experiencia de compra. Estos trabajos, demuestran que los procesos de *Sentiment Analysis* están en un momento de evolución en la medida que se va comprendiendo cada vez más lo que hace que los clientes se motiven a optar por una empresa comercial u otra. De esta manera las organizaciones pueden identificar puntualmente qué atributos de sus productos o servicios valoran más sus clientes y, en consecuencia, atender las debilidades que puedan tener para luego promover mediante campañas de mercadeo las prestaciones que brindan. Para su estudio los autores analizan los *user-generated contents (UGC's)* de la página *booking.com*; utilizan *LSA* para describir una matriz con las evaluaciones positivas y las negativas. Recolectan las opiniones de clientes de hoteles de las 100 ciudades más grandes de EEUU.

## 2.3 Procesos generales de Web Mining

El reconocimiento de las oportunidades que puede ofrecer *Big Data* en el entendimiento de patrones del comportamiento humano y fenómenos sociales han ofrecido resultados no detectados con enfoques de análisis anteriormente conocidos. Cada vez más las organizaciones se empeñan en explotar recursos de datos administrativos con el objetivo de aprovechar al máximo lo que pueda ayudarles a comprender más el comportamiento de sus clientes. El análisis de los datos administrativos resulta complejo para los analistas, debido a que la información que se recolecta no se hace con el fin de ser explotada para fines de investigación. Es por esto que la combinación de las más reconocidas técnicas y metodologías de gestión que contribuyan al desarrollo de proyectos de *Text Mining* es esencial para lograr el éxito en compromisos como estos.

Tal como Bandyopadhyay y Pal (2007) expresan, podemos resumir que las tareas principales de los procesos de *Web Mining* son los siguientes:

- *Information Retrieval (IR)*: Que comprende la recolección de documentos a través de la web, descartando los considerados como irrelevantes. Tal como funcionan las búsquedas de Google y las funciones de *WebCrawlers*.
- *Information Extraction (IE)*: Se ocupa de extraer conocimiento automáticamente (sin intervención humana) una vez que los documentos han sido recolectados.
- *Generalization*: En esta fase se aprenden conceptos generales a partir de los documentos dados. Es aquí en que se aplican las técnicas de reconocimiento de patrones y *Machine Learning*, como *Classification*, *Clustering* y *Association Rules Mining* para obtener informaciones relevantes.
- *Analysis*: Se ocupa de las tareas de entendimiento, visualización e interpretación de los patrones una vez descubiertos en la fase *Generalization*.

## 2.4 Paradigma tradicional y reciente sobre el análisis de los datos

A continuación se presentan cuatro paradigmas en *Big Data* relacionados a la ciencia del conocimiento y estudio del comportamiento social como parte de la combinación de procesos tradicionales y la adopción de nuevas técnicas para la gestión de los datos, (Mahmoodi et al., 2017). En su análisis, los autores resaltan la comparación y convivencia de las siguientes visiones:

### Predicción y Explicación

Siendo la Predicción el modelo para pronosticar futuros comportamientos a partir de estudio de la historia de la información disponible, mientras que la Explicación, más orientada a procesos tradicionales, no es descartada en los procesos de *Big Data* y puede contribuir a mejores modelos de predicciones.

### Enfoques Inductivos y Deductivos

En los modelos de trabajo tradicionales Deductivos (impulsados por teorías) regularmente se inicia la investigación científica de los datos a partir de hipótesis no fundamentadas y el equipo de trabajo se ocupa en demostrar la veracidad de la motivación del proyecto. Esta manera tradicional de abordar tareas de análisis puede descuidar todo lo que se podría descubrir a partir de los datos disponibles. En el enfoque Inductivo (impulsado por datos), se puede aprender a partir de los datos que se van observando, permitiendo a los investigadores deducir teorías a partir de las informaciones recolectadas y analizadas.

## Grandeza y Representatividad

Son otros paradigmas que comparan los autores como enfoques modernos y tradicionales. La Representatividad refiere al muestreo de datos para fines de estudio mientras que la Grandeza o (*Big Data* por definición) se ocupa de aprovechar la vasta cantidad de datos con el objetivo de descubrir patrones que anteriormente no eran posibles por no ser fácilmente accesibles con la tecnología disponible.

## Acceso de Datos versus Independencia del científico

La cuestión del acceso a los datos a expensas de la independencia científica y los conflictos de intereses puede ser complicado en el enfoque de *Big Data*. Los intereses y las prácticas de las empresas pueden sesgar la investigación y restringir las informaciones que ponen a disposición de otros. La independencia científica solo puede ser sustentada si todas las partes se subscriben al estándar determinado para el análisis de los datos.

## 2.5 Metodologías de gestión de proyectos utilizadas actualmente en Minería de Datos

La complejidad asociada en aplicar las técnicas correctas en procesos relacionados a extraer informaciones útiles a partir de grandes volúmenes de datos hace que las organizaciones emprendan proyectos con alto grado de incertidumbre debido a que no existen metodologías claramente predefinidas que cubran todas las problemáticas y provean soluciones que envuelven estos tipos de trabajos de análisis, diseño de solución, desarrollo de software e implementación de sistemas cumpliendo con las principales restricciones que tienen la mayoría de los proyectos: Tiempo, Costo, Alcance y Calidad. Diversas metodologías de Dirección de Proyectos han sido estudiadas en la literatura y concebidas como estándar para proyectos frecuentemente abordados en áreas como la Construcción, de ámbito Social, Desarrollo de Software, entre otros. Sin embargo, los proyectos relacionados a Minería de Datos que requieren personal con perfiles cualificados, uso de tecnologías de vanguardia, gestión efectiva de entregables con poca claridad en requerimientos iniciales, trabajo en armonía de diversas áreas de conocimiento y control de cambios de alto impacto durante el desarrollo de la solución, requieren ser trabajados en un marco distinto por la naturaleza de las tareas que se llevan a cabo.

Como metodologías en proyectos y ejecución de procesos altamente predecibles, ha proliferado el uso de Lean por su enfoque en la optimización de los recursos y mejora continua. En otro orden, proyectos tradicionales de Desarrollo de Software suelen basarse en las mejores prácticas descritas en el PMBOK, Prince2, Norma ISO 21500, entre otras. A partir de 2001, surgen formalmente las metodologías ágiles que se han estado aplicando en proyectos de desarrollo software con cierto grado de incertidumbre y poca definición inicial de los requerimientos. Los objetivos y principios de las metodologías ágiles han sido adoptados también en proyectos de *Business Intelligence* (BI) y *Data Science* por la flexibilidad que representa su ciclo de vida iterativo de desarrollo del proyecto y promoción de la colaboración en equipo, (Larson y Chang, 2016). El uso de metodologías ágiles representa un enfoque menos formal, más dinámico y centrado en el cliente. Procesos de *Machine Learning* dependen en gran medida de procesos típicos de las metodologías ágiles en las que frecuentes entregas a los *stakeholders* son requeridas. Un cambio en los requerimientos del proyecto significa un cambio de alcance, lo cual impacta en tiempo, recursos y presupuestos (los aspectos fundamentales en la gestión de proyectos). El enfoque tradicional de gestión de proyectos obedece a seguir el plan y descartar los cambios. Las metodologías ágiles promueven eliminar la burocracia, y se enfoca más bien en entregas tangibles frecuentes que pueden servir para medir el curso que toma el proyecto.

Conceptos de la metodología ágil *Scrum* que se han utilizado en BI son: *User Stories*, *Sprint Backlogs*, *Product Backlogs*, *Sprints* y *DailyScrums*. Los requerimientos de BI son divididos en pequeños relatos que luego son empaquetados en una colección de narraciones para abordar el proyecto. Cada relato es designado, desarrollado, probado y liberado. Un *Sprint* dura de una a dos semanas y contiene un ciclo de requerimientos, análisis, diseño, desarrollo y prueba de usuarios finales. Los relatos pueden ser agrupados en un *Sprint Backlog* o en un *Product Backlog*. Los *Sprint backlog* se refieren al trabajo que el equipo de desarrolladores completa durante el *Sprint*, mientras que el *Product Backlog* es la lista de relatos ordenados por prioridad para ser considerados en el próximo *Sprint*. Los usuarios están involucrados en todos los *Sprints*. Reuniones diarias de 15 minutos o menos para la verificación del estatus del proyecto son realizadas cada día, (Muntean y Surcel, 2013).

Las metodologías ágiles resultan ser hoy día las más adecuadas para el desarrollo de proyectos de *Data Science*, sin embargo hay retos aun que deben asumirse y procesos a mejorar para una efectiva gestión debido a que estas metodologías no han sido concebidas específicamente para proyectos con estas características. La naturaleza de estos proyectos induce frecuentemente a que las organizaciones excedan el presupuesto destinado para los fines propuestos, provoca que la



planificación se vea afectada recurrentemente, el alcance se comprometa, no se reconozca el beneficio real del proyecto, y que los usuarios finales y la organización misma no esté a gusto con los resultados si no son correctamente gestionados.

## 2.6 Evaluación de metodologías alternas para gestión de proyectos

El uso de metodologías ágiles es un cambio importante en la forma en que las organizaciones administran los proyectos si se compara con técnicas tradicionales de Cascada. A parte de la entrega a tiempo de productos de softwares, la calidad de un proyecto implica la satisfacción de los involucrados y la fuerza laboral. Si la fuerza laboral está satisfecha en su entorno de trabajo, será más productiva, creativa e innovadora, y por tanto será capaz de crear soluciones más robustas. Las metodologías existentes generalmente se enfocan en los factores de éxito administrativos y técnicos de los proyectos de desarrollo de software, como entrega oportuna y conservación del presupuesto designado. La mayoría de las metodologías utilizadas en proyectos de desarrollo de software se basan en técnicas matemáticas y estadísticas, no sobre el factor personas. A menudo, los factores de éxito conductuales y psicológicos no son tomados en cuenta en la gestión de proyectos. Bien es sabido que en general la productividad y el éxito de los proyectos dependen en gran medida del bienestar psicológico de los individuos y del grupo en su conjunto. En la literatura, Kashyap (2017) propone para el desarrollo de software la integración entre las metodologías ágiles y la inteligencia holística de los integrantes del equipo. El modelo propuesto puede ser provechoso para entender a las personas difíciles y corregir sus comportamientos. La meta es alinearlos a la visión, misión y propósito de la organización.

Si bien es cierto que las metodologías ágiles se adaptan a las necesidades operativas de proyectos de *Data Science*, no presentan detalles específicos asociados a estos tipos de proyectos. Otro tipo de metodologías popularmente conocidas presentan un modelo detallado de procesos asociados a esta clase de proyectos, como expresan Cios y Kurgan (2005), y Sharma y Osei-Bryson (2009) respecto a la metodología CRISP-DM (CRoss-IndustryStandardProcessforDM) y su guía que propone (Chapman et al., 2000). Cabe destacar que esta metodología carece de un flujo de trabajo y descripción de cómo debe implementarse en una organización. Los autores de los artículos reconocen la importancia de la metodología CRISP-DM para determinar los procesos dependientes de la Fase de Entendimiento del Negocio; esencial para el éxito de un proyecto. La metodología CRISP-DM es objeto de estudio por Marbán et

al., (2009), en el cual la comparan con otras metodologías maduras en el desarrollo de software, lo cual la hacen ver incompleta por la falta de importantes procesos que serían efectivos en proyectos de este tipo. Por tal razón los autores recomiendan una contraparte Agile de CRISP-DM.

En otro orden, como en cualquier proyecto convencional, es fundamental realizar las estimaciones de costos normalmente desglosando la lista de actividades y recursos asociados. No obstante, en proyectos de *Data Science* es algo distinto por su naturaleza no tan predecible. Marbán et al., (2008) proponen un modelo pensado para proyectos de esta índole, en el cual luego de estudiar el método paramétrico COCOMO para estimar costos basándose en ‘Cost Drivers’ específicos a proyectos de *Data Science* y comparándolo con presupuestos reales de proyectos utilizando regresión lineal; el modelo es DMCoMo y presenta estimaciones cercanas a la realidad. Considerando una mejora en el modelo de estimación de costos DMCoMo se debe tener en cuenta una optimización enfocada a pequeñas y medianas organizaciones, (Pytel et al., 2012).

Como aspectos importantes a considerar en los procesos de gestión, primordialmente los relacionados a desarrollo de software, son la planificación, evaluación, respuesta y control de los riesgos. La gestión del riesgo es un compromiso complejo del cual depende en gran medida el éxito de los proyectos. Considerando la alta tasa de proyectos fallidos en el área de software, lo cual demuestra *The Standish Group* en su reporte “*CHAOS Report 2016*” que solo el 29% de los proyectos estudiados de desarrollo de software resultaron ser exitosos durante el citado año, es de suma importancia enfocarse en una gestión efectiva de los riesgos. El uso de técnicas para monitorear y controlar los riesgos debe imperar para no afectar el curso aceptable que permitan los proyectos. Dentro de los procesos de gestión, para las tareas de planificación y análisis se puede aprovechar el marco de trabajo que proponen investigadores en la literatura como complemento a la vasta cantidad de herramientas disponibles, y que normalmente no son aprovechadas en los equipos de proyectos, (Hu et al., 2013).

## 2.7 Propuesta de trabajo

La actual economía aumenta la demanda de soluciones rentables que permitan a las organizaciones ganar ventaja en la competitividad empresarial. Por esta razón, más y más compañías buscan metodologías efectivas de gestión que les permita mejorar sus productos y/o servicios, perfeccionar sus procesos, disminuir los costos, mejorar la rentabilidad y lograr la satisfacción de sus clientes. No todos los modelos y metodologías conocidas en el área de desarrollo de software aplican para proyectos de *Data Science*, combinaciones de modelos, metodologías, técnicas y herramientas deben utilizarse debido a sus características especiales. El uso de metodologías Agile y aprovechamiento de las características de CRISP-DM pueden contribuir a gestionar proyectos de Data Science cumpliendo con las principales restricciones que implican abordar un proyecto. El objetivo de este trabajo consiste en proponer un marco de trabajo integral para dirección de proyectos de *Text Mining* con miras a captación de nuevos clientes tomando en cuenta las diversas técnicas y herramientas que se encuentran en plena evolución, las bondades que ofrecen las metodologías más utilizadas según la literatura con las cuales se podría controlar las restricciones de tiempo y alcance, las recomendaciones avaladas para gestionar los procesos críticos de costos y riesgos, los conocimientos aprendidos en clases y la experiencia en el ámbito laboral.

# Capítulo 3

## Metodología de la investigación

“La gestión de proyectos es el arte y la ciencia de convertir la visión en realidad”  
Dr. Turner

### 3. Metodología de la investigación

La investigación aplicada en este trabajo es de tipo No Experimental de carácter Exploratorio, en la que se evalúan los estándares avalados por la comunidad profesional del PMI, los artículos científicos relacionados a las mejores prácticas relacionadas a minería de datos y la bibliografía correspondiente al máster interuniversitario de Dirección de Proyectos de la Universidad de La Rioja, la Universidad Pública de Navarra y la Universidad de Oviedo.

Como parte del alcance de la investigación, no se estarán aplicando mediciones de campo que impliquen evaluar la efectividad de los procesos en las etapas de los proyectos de minería de datos que llevan a cabo las organizaciones para captar nuevos clientes. Se evaluarán los diversos enfoques de dirección de proyectos más apropiados al área de estudio ya que no se describen metodologías universalmente aceptadas en la literatura para los fines descritos.

Los siguientes pasos han de ser efectuados con el objetivo de llevar a cabo la evaluación de las prácticas conocidas en la literatura y permitir la elaboración de una propuesta de marco de trabajo apropiado que beneficie las tareas de dirección de proyectos que permitan ampliar la cartera de clientes en las empresas.

#### **Etapas 1: Clasificación de los modelos de Minería de Datos y delimitación del estudio**

En esta etapa se describen las dos principales ramas de técnicas para el análisis de los datos. Por un lado, se presentan las herramientas más tradicionales para la exploración de los datos y explicación de los comportamientos pasados, de igual modo se explica cuáles son los principales modelos de predicción.

Las etapas siguientes obedecen a la evaluación de prácticas de dirección de proyectos que utilizan los modelos de predicción a partir de patrones de comportamiento de los clientes identificados en los datos.

#### **Etapas 2: Evaluación de modelo de procesos CRISP-DM**

Se evalúa el estándar "de facto" para el desarrollo de proyectos de Minería de Datos conocido como CRISP-DM. En esta etapa se presentan cuáles son sus fases y tareas asociadas. Luego de analizar su aplicabilidad para proyectos de creación de modelos de predicción se muestran sus ventajas y desventajas.

### **Etapas 3: Uso de metodologías prescriptivas**

Se estudia la posibilidad de utilizar metodologías prescriptivas tradicionales en lugar de basarse en el modelo de procesos CRISP-DM. Se analizan las etapas de Requerimientos, Diseño, Implementación, Verificación y Mantenimiento.

### **Etapas 4: Evaluación de guía de conocimientos del PMBOK**

En esta etapa se estudian las recomendaciones que ofrece el Project Management Institute en la guía de conocimientos PMBOK para la dirección general de proyectos con enfoque a Minería de Datos y aplicación de técnicas para construcción de modelos predictivos.

Se estudia el ciclo de vida que propone CRISP-DM para la dirección de los proyectos. Se evalúan las fases que las componen y sus puntos de revisión.

Se realiza un análisis sobre CRISP-DM para identificar las tareas de dirección de proyectos que incorpora en su modelo de procesos.

Luego, se presentan conceptos de la guía de conocimientos del PMBOK sobre los procesos de dirección y sus correspondientes áreas de conocimientos.

### **Etapas 5: Aplicabilidad procesos del PMBOK en CRISP-DM**

A partir de los procesos del PMBOK se analiza su aplicabilidad en la dirección de proyectos de Minería de Datos y a la vez se homologan a las etapas-tareas que componen el modelo de procesos CRISP-DM.

### **Etapas 6: Uso de metodologías ágiles**

Se realiza un estudio sobre los principios de las metodologías ágiles y su posible relación con el enfoque de CRISP-DM.

### **Etapas 7: Marco de trabajo Scrum**

Se describen los conceptos del marco de trabajo Scrum que pueden servir para la adopción de un enfoque ágil en la dirección de proyectos de Minería de Datos.

### **Etapas 8: Aplicabilidad actividades de Scrum en CRISP-DM**

A partir de las actividades del marco de trabajo Scrum se analiza su aplicabilidad en la dirección de proyectos de Minería de Datos y a la vez se homologan a las etapas-tareas que componen el modelo de procesos CRISP-DM.

### **Etapas 9: Propuesta Marco de Trabajo**

A partir de los modelos de trabajo estudiados, se ofrece una propuesta que incorpore las tareas de CRISP-DM orientadas al desarrollo del producto, las recomendaciones de la guía de conocimientos del PMBOK aplicables a proyectos de Minería de Datos y el enfoque ágil de Scrum. Se consideran las herramientas vistas en los cursos del programa de maestría que contribuyen a la gestión efectiva de los proyectos en estudio.

Se ofrece un marco híbrido que conjugue las prácticas idóneas que ofrecen cada una de las fuentes estudiadas.

# Capítulo 4

## Resultados y discusión

“La adopción de metodologías ágiles no es un cambio de proceso; es un cambio de cultura”



## 4. Resultados y discusión

Los **Datos** representan para una organización la materia prima a partir de la cual se puede iniciar la exploración para comprender su significado.

Al momento en que los Datos tienen un significado para la empresa se convierten en **Información**.

Cuando los analistas pueden construir un modelo a partir de las **Informaciones**, se genera un valor agregado, el cual podemos llamar **Conocimiento**. Es esto lo que persiguen las empresas para implementar las estrategias de negocios que permitirán su crecimiento en el mercado.

Para obtener **Conocimiento** a partir de los **Datos** es preciso abordar proyectos de Minería de Datos.

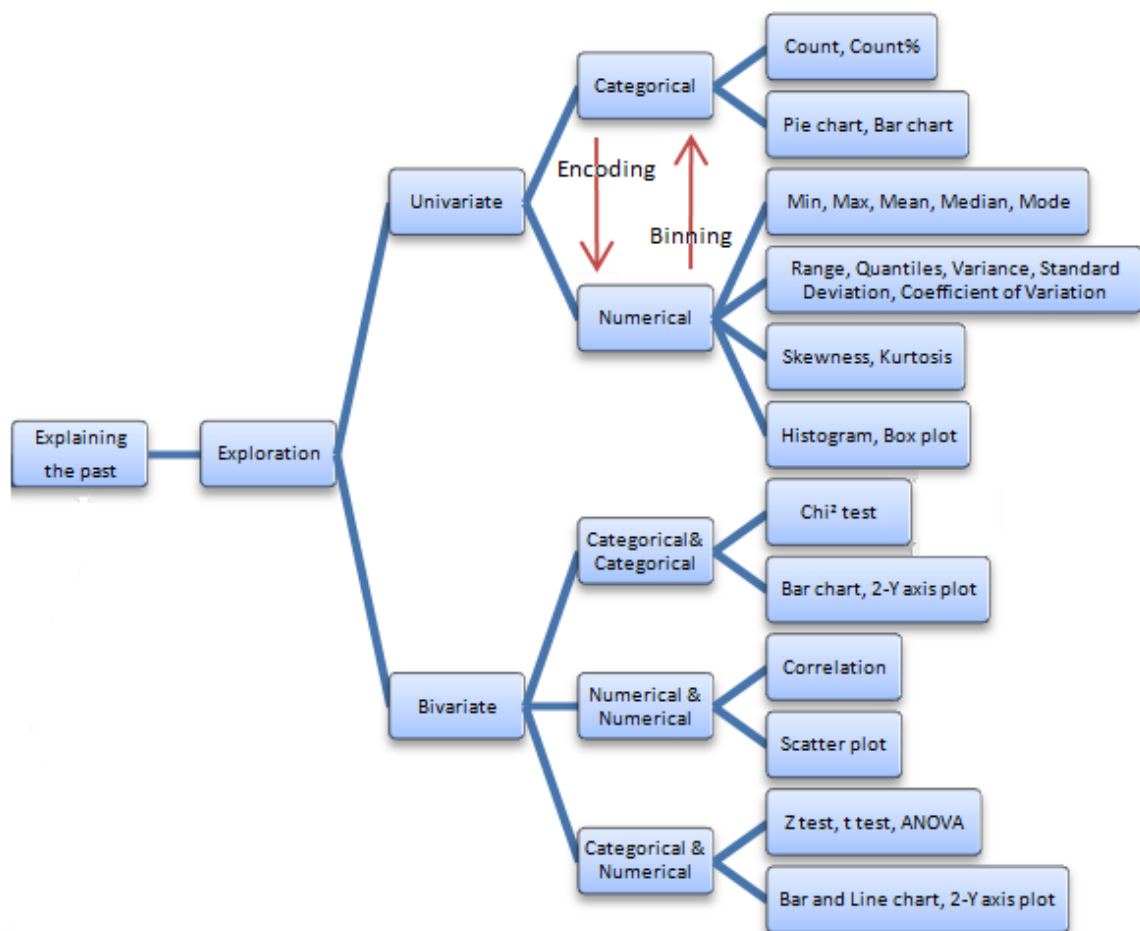
La Minería de Datos se puede describir como el conjunto de técnicas y tecnologías que nos permiten explorar grandes volúmenes de datos de manera automática o semiautomática. El principal objetivo de aplicar estas técnicas es detectar patrones repetitivos y tendencias que nos expliquen el comportamiento de los datos en determinado contexto.

Los proyectos de Minería de Datos se caracterizan por la incertidumbre que los envuelve. Pocos o ambiguos requerimientos, numerables herramientas y técnicas disponibles y un determinado propósito que lograr.

## 4.1 Técnicas de Minería de Datos

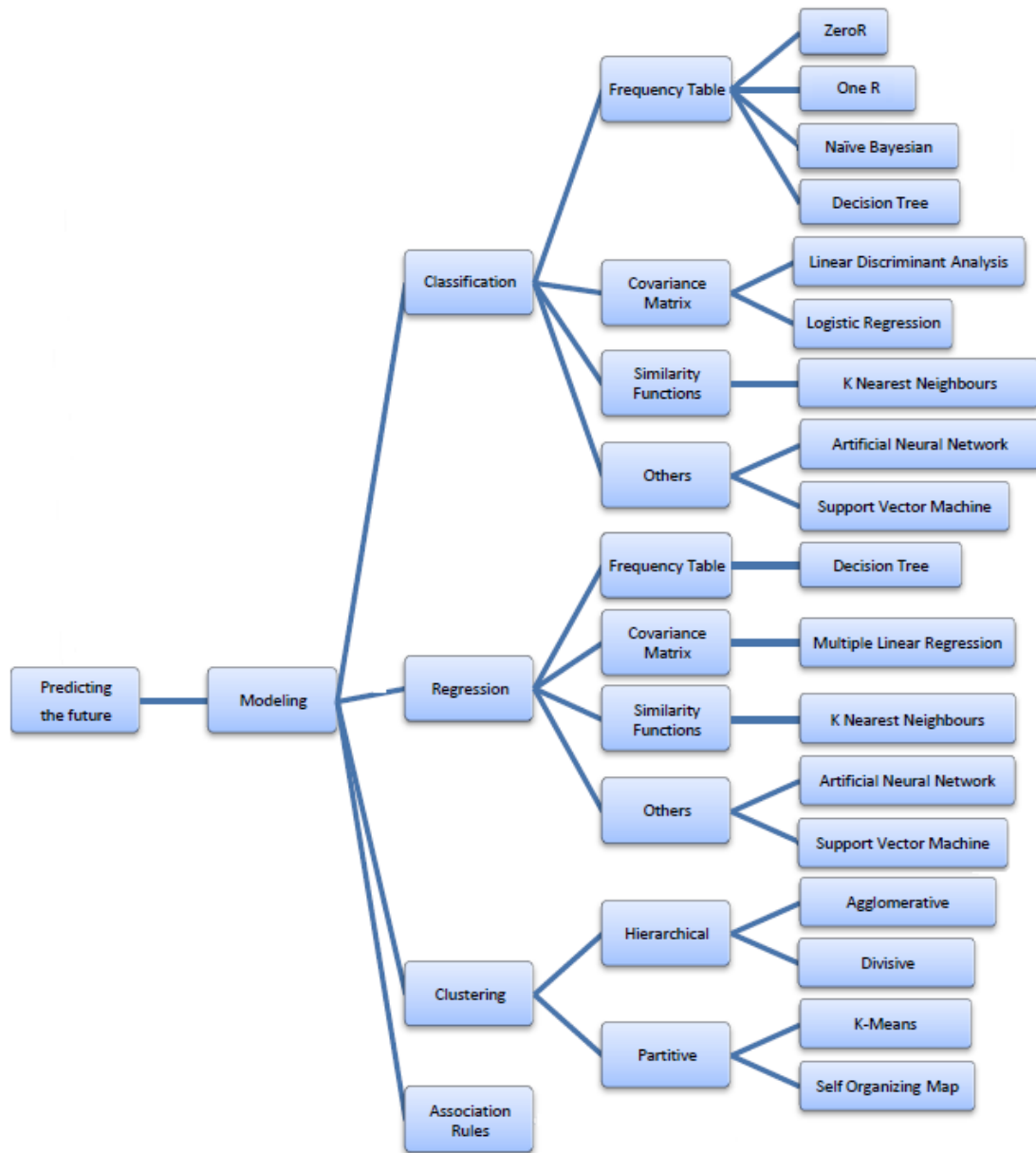
Diversas técnicas son utilizadas en Minería de Datos con el fin de comprender las informaciones que residen en los sistemas corporativos de las Organizaciones.

A medida que transcurre el tiempo aparecen nuevas herramientas para la interpretación de los datos. Por un lado, contamos con herramientas de exploración para fines de explicar el pasado; detectar patrones en base a la exploración de los datos disponibles. Los científicos de datos pueden determinar que significan las relaciones en las informaciones que descubren a partir de los análisis estadísticos que llevan a cabo.



*Ilustración 1. Técnicas de Minería de Datos para explicar el pasado (Sayad, s.f.)*

En la Ilustración 1 podemos observar la clasificación de las técnicas más tradicionales de análisis estadísticos para la comprensión de los datos en manera de explicación del pasado.



*Ilustración 2. Técnicas de Minería de Datos para predecir futuros comportamientos (Sayad, s.f.)*

En la Ilustración 2 podemos observar la clasificación de las técnicas más tradicionales de modelado para la comprensión de los datos en manera de predicción del futuro.

Este estudio se enfoca metodologías de dirección de proyectos basados en análisis de predicción a futuro, en los cuales se utilizan modelos matemáticos para comprender mejor las variables que determinan el éxito del caso de estudio. Los análisis predictivos se componen de fórmulas que comparan los sucesos exitosos y fallidos en el pasado, y a partir de cálculos con dichas fórmulas se determina un patrón de comportamiento que describe acontecimientos futuros.

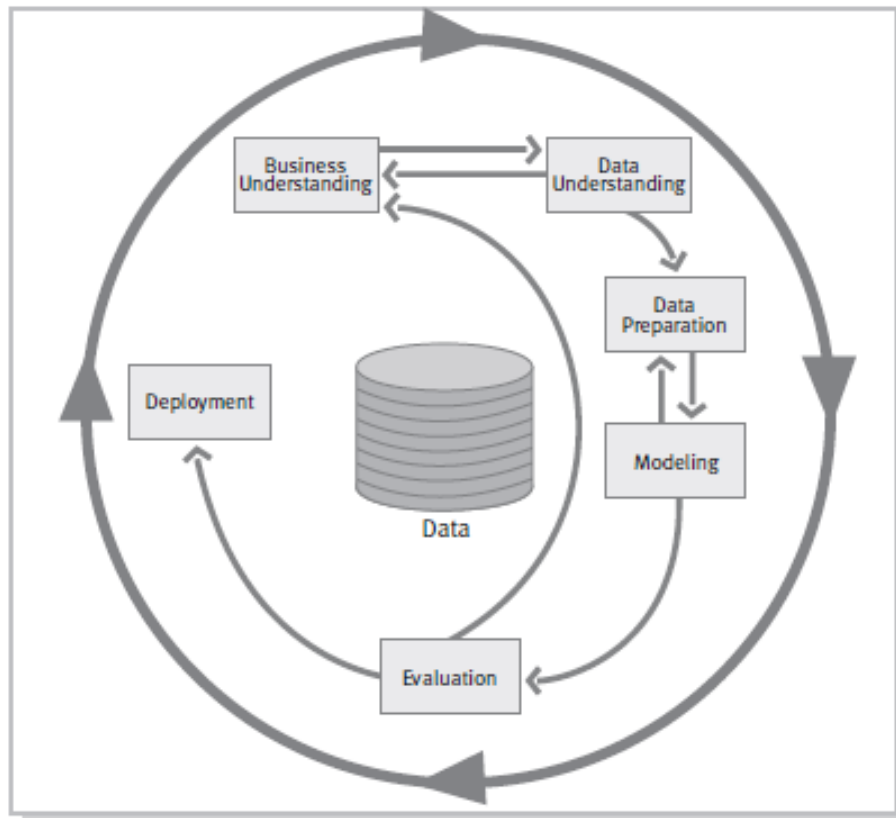
## 4.2 CRISP-DM

CRISP-DM se crea a finales del año 1996. Concebido como un proyecto de la Unión Europea liderado por cinco organizaciones: SPSS, Teradata, Daimler AG, NCR Corporation y la empresa de seguros OHRA.

En la actualidad, el modelo de procesos CRISP-DM es el más utilizado y efectivo para la ejecución de proyectos de Minería de Datos, sin embargo, carece en parte, de prácticas ya conocidas en la industria de Dirección de Proyectos que contribuyen a las organizaciones dar un seguimiento efectivo sobre la viabilidad de sus emprendimientos.

CRISP-DM abstrae en seis fases las actividades que se realizan en un proyecto de Minería de Datos. Este modelo incluye además de sus tareas, los entregables para cada una de ellas. Presenta un ciclo de vida cuyo flujo no debe ser estrictamente en el orden que se plantea, lo cual brinda flexibilidad a la hora de desarrollar los modelos necesarios para la consecución de los objetivos del proyecto.

A continuación, se presenta un gráfico que ilustra las fases que lo componen.



*Ilustración 3. Metodología CRISP-DM, (Chapman et al., 2000)*

Las fases que conforman el modelo de procesos CRISP-DM obedecen a las necesidades que han determinado sus autores en proyectos generales para extraer conocimiento de los datos. Se describe a continuación en qué consisten:

### Entendimiento del negocio

Es la más importante de las fases del proyecto ya que es acá donde se identifican los requerimientos y objetivos, los cuales muchas veces ni siquiera el mismo cliente tiene detalle de ellos. Acá es donde se define en términos medurables (que puedan ser medidos), los objetivos y criterios de éxito del proyecto y para tal fin, el líder o líderes del proyecto deben tener un conocimiento suficiente y bien estructurado de las funciones de negocio de la empresa.

#### Tareas:

- **Determinar los objetivos de Negocio:** Tarea en la cual se identifica el objetivo principal desde la perspectiva del Negocio que desea lograr el cliente.

- **Evaluar la situación:** Esta tarea está destinada a identificar los recursos necesarios para lograr los objetivos planteados por el cliente, explorar las asunciones de los involucrados y otros factores que puedan incidir para lograr los objetivos planteados o comprometer el plan de proyecto.
- **Determinar metas de Minería de Datos:** Determinar las metas técnicas de Minería de Datos que contribuyan a lograr los objetivos de Negocio.
- **Crear el Plan de Proyecto:** En esta tarea se define un plan de acciones necesarias para alcanzar las metas de Minería de Datos, y por consiguiente lograr los objetivos de Negocio. Este plan ha de definir los pasos a seguir para el resto del proyecto, incluyendo la selección de herramientas y técnicas a utilizar.

## Entendimiento de los datos

Identificar cómo y dónde reposa la información con que cuenta la organización para llevar a cabo el proyecto. Ubicación de los servidores, los permisos de acceso, el formato en que se encuentran los datos (Oracle, SQL, Excel y demás).

### Tareas:

- **Recoger datos iniciales:** Obtener o tener acceso a los datos determinados en el listado de recursos.
- **Describir los datos:** Examinar en detalle la información disponible para abordar el proyecto.
- **Explorar los datos:** Explorar los datos y analizar su potencial explotación vía técnicas de Minería de Datos. Validar que la información disponible permite lograr las metas de Minería de Datos.
- **Verificar la calidad de los datos:** Examinar la calidad de los datos, respondiendo a las preguntas: ¿Están completos los datos para los modelos a construir?, ¿Contienen errores los datos?, Y si ¿Son despreciables los errores o son significativos?, ¿Falta información?; de ser así, como se pueden representar, donde y que tanta.

## Preparación de los datos

Es el paso que generalmente más tiempo toma ya que, en la mayoría de los casos, los *datamarts* o *data warehouses* no han sido diseñados para realizar proyectos de Minería de Datos, además se cuenta con información dispersa y data no estructurada.

### Tareas:

- **Seleccionar los datos:** Decidir cuáles son las informaciones a utilizar para el análisis.
- **Limpiar los datos:** Mejorar la calidad de los datos al nivel requerido de las técnicas de análisis.
- **Construir los datos:** En esta tarea se consideran operaciones de preparación de datos para ofrecer sentido a información que pueda existir incompleta o se requiera nuevos datos a partir de los existentes.
- **Integrar los datos:** Los datos son combinados a partir de diversas tablas o registros para crear nuevos registros o valores.
- **Formatear los datos:** Modificaciones en los datos que no cambian su significado, ya que así se requieren para las herramientas de modelado.

## Creación de los modelos

Una vez se ha entendido el negocio y los datos, y ya éstos se encuentran listos para trabajar con ellos, llega la hora de elegir el modelo numérico/predictivo que mejor respuesta de a los objetivos del proyecto.

### Tareas:

- **Seleccionar la técnica de los modelos:** Seleccionar la técnica de modelado a ser utilizada para lograr las metas de Minería de Datos si aún no ha sido determinada en la fase de Entendimiento del Negocio. Determinar de forma específica las técnicas a utilizar, árboles de decisiones, redes neurales, entre otras.



- **Generar diseño de pruebas:** Elaborar diseños que permitan validar la calidad y efectividad de los modelos a construir.
- **Construir el modelo:** Utilizar en los datos disponibles las herramientas de modelado para elaborar nuevos modelos.
- **Evaluar el modelo:** Interpretar los modelos. Contrastar con los criterios de éxito y los diseños de pruebas.

## Evaluación y pruebas

Los modelos seleccionados deben ser validados y re-entrenados para verificar su precisión y eficacia.

### Tareas:

- **Evaluar los resultados:** Los pasos previos evalúan la precisión y lo concerniente a los modelos. Este paso evalúa el grado en el cual los modelos cumplen con los objetivos de Negocio y busca identificar las razones por las cuales no sería efectivo.
- **Revisar los procesos:** Determinar si algo ha sido pasado por alto en la construcción y evaluación de los modelos.
- **Determinar próximos pasos:** Dependiendo de los resultados obtenidos en la evaluación de los modelos, el equipo decide cómo proceder. Se toma la decisión de terminar el proyecto, desplegar cambios, iniciar nuevas iteraciones o programar nuevos proyectos.

## Despliegue o puesta en producción

El último paso donde finalmente se empieza a sacar provecho de los modelos para resolver los problemas de negocio detectados en la etapa inicial.

### Tareas:

- **Planificar despliegue:** En esta tarea se toman los resultados de la evaluación y se determina una estrategia de despliegue.

- **Planificar monitoreo y mantenimiento:** Planificación de monitoreo y mantenimiento si los resultados de los procesos de minería pasan a formar parte del día a día de las actividades de Negocio y su ambiente productivo.
- **Elaborar reporte final:** Al final del proyecto, el equipo prepara un reporte final que refleje el resumen de lo ejecutado, y las experiencias del proceso. Puede ser también una presentación comprensiva de los resultados obtenidos.
- **Revisar el proyecto:** Evalúa las situaciones positivas y negativas que surgieron durante el proyecto, y que podría mejorar.

## Ventajas y Desventajas de CRISP-DM

### Ventajas

- Es un modelo simplificado que describe las principales actividades que son necesarias para desarrollar un producto de Minería de Datos.
- Es el modelo más conocido en la industria, y contará de fácil aceptación por el equipo y la organización.

### Desventajas

- No considera procesos esenciales para la gestión efectiva de los proyectos.
- Es reconocido como una metodología de dirección de proyectos, sin embargo, no se detalla la manera de cómo llevar a cabo todas las actividades, más bien se enfoca en lo que se debe hacer. Esto lo convierte en un modelo de procesos.
- El ciclo de vida no fomenta la armonía entre equipos de trabajo de diversas áreas.
- Presenta un ciclo de vida Predictivo con retroalimentación entre fases. La guía establece que es un ciclo de vida Iterativo, no obstante, la consecución de sus fases tiene el enfoque de elaborar un producto final al concluir las etapas. Un modelo Predictivo en proyectos de Minería de Datos, en los cuales

prevalece la incertidumbre de las necesidades del Negocio, resulta no ser un modelo ideal según se demuestra en este estudio.

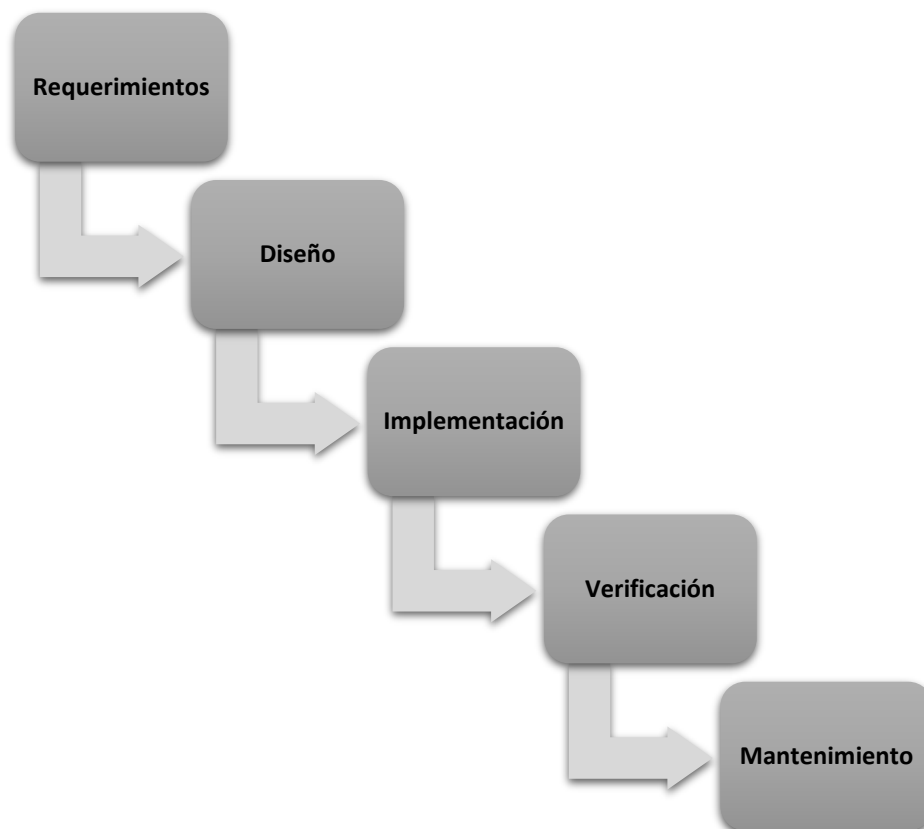
- Es un modelo de procesos que no tiene soporte por la industria. Una vez liberada la guía, no ha tenido actualizaciones que contribuyan a optimizar los procesos que deban considerarse.

## 4.3 Metodologías Predictivas

Las metodologías predictivas (en Cascada) tienen metas para cada fase del desarrollo del proyecto. Una vez una fase se completa, el proyecto se mueve a otra y no puede retroceder una vez ha pasado por ella, (Crisp y McKenna, 2016).

El modelo en Cascada define un método linear, cerrado y secuencial.

La **Ilustración 4** muestra las fases generales de un ciclo de vida en Cascada o Predictivo.



*Ilustración 4. Ciclo de vida en Cascada*

### Requerimientos

Fase en la que el analista de negocios define los requerimientos. Se establece cual será el producto final a construir.

Esta fase coincide en cierta medida con el enfoque de la etapa **Entendimiento del Negocio** de CRISP-DM.

## Diseño

En esta fase el equipo describe como debe ser construido el producto final. Técnicas y herramientas a utilizar son seleccionadas. Se establecen acuerdos respecto a los criterios de aceptación. El resultado de esta fase ha de servir para la fase de Implementación en la que los programadores construirán el sistema solicitado.

CRISP-DM no considera una etapa de diseño como tal, ya que por la naturaleza de sus actividades no se requiere. Tareas específicas relacionadas a la comprensión de los datos son consideradas en las fases **Entendimiento de los Datos** y **Preparación de los Datos**.

## Implementación

Luego de contar con el diseño, continua la fase de Implementación en la cual se realiza el trabajo necesario para construir el producto objetivo.

En CRISP-DM se construye el producto objetivo en la fase **Creación de los modelos**. Es aquí en la que los expertos dedican su tiempo al desarrollo de la solución.

## Verificación

En este momento se procede a validar la calidad del producto. Se valida que el producto cumple con las expectativas del cliente. Se ejecutan planes de pruebas preparados en la fase de Diseño.

CRISP-DM cuenta con la fase **Evaluación y pruebas** para los fines de validar los modelos construidos.

## Mantenimiento

Una vez valido el producto, se procede con el despliegue de los cambios para uso por parte del cliente. Se da seguimiento al funcionamiento del sistema de manera tal que, si es necesario corregir o mejorar alguna característica del producto ya entregado, se procede a evaluar si se puede considerar el cambio e iniciar nuevamente el ciclo hasta realizar otro despliegue.

La fase **Despliegue o puesta en producción** existe en CRISP-DM con el objetivo de realizar los despliegues necesarios y realizar las tareas de finalización del proyecto.

## Evaluación de las metodologías predictivas

Por lo que se puede apreciar, CRISP-DM es un modelo de procesos creado a partir de un enfoque predictivo que permite cierta flexibilidad retrocediendo entre una fase y otra.

Debido a la alta interacción entre equipos y flexibilidad que requieren los proyectos de Minería de Datos, una metodología en Cascada no representa una alternativa factible sobre la cual crear soluciones impredecibles de captación de clientes vía examinación de datos y creación de modelos predictivos.

## 4.4 Recomendaciones de la guía del PMBOK con enfoque a proyectos de Minería de Datos

La dirección de proyectos es la aplicación de conocimientos, habilidades, herramientas y técnicas a las actividades del proyecto para cumplir con los requisitos del mismo. Se logra mediante la aplicación e integración adecuadas de los procesos de dirección de proyectos identificados para el proyecto, (Project Management Institute, 2017a).

Un concepto erróneo existe sobre la aplicabilidad en la gestión de proyectos de desarrollo de software y entornos ágiles las prácticas descritas en la guía del PMBOK. Se evalúa en este capítulo las recomendaciones que ofrece el Project Management Institute en su guía de conocimientos, (Sliger y Consulting, 2008).

La implementación de prácticas reconocidas sobre gestión de proyectos puede beneficiar a las organizaciones en el emprendimiento de trabajos de Minería de Datos para los cuales no existe un estándar adoptado universalmente. Se evalúa a continuación la aplicabilidad del enfoque del PMI a la dirección de estos tipos de proyectos.

### Ciclo de vida del proyecto

El ciclo de vida es la serie de fases que atraviesa un proyecto desde su inicio hasta su conclusión. CRISP-DM describe seis fases compuestas de las principales actividades necesarias para completar un proyecto de Minería de Datos. Procesos básicos para la gestión integral del proyecto deben ser considerados en adición a los expuestos en el modelo de procesos en estudio.

Las actividades que componen esta metodología describen un ciclo de vida predictivo. Los estudios realizados y nuestra experiencia de trabajo demuestran que estos tipos de proyectos han de ser desarrollados en un ambiente colaborativo con enfoque adaptativo.

### Fase del proyecto

Conjunto de actividades del proyecto relacionadas lógicamente que culmina con la finalización de uno o más entregables.

CRISP-DM considera en cada fase las actividades necesarias para cumplir con los objetivos de un proyecto de Minería de Datos.

La propuesta resultante de este estudio debe presentar una serie de fases entendibles para la organización y el equipo de trabajo. Debe ser un marco de trabajo que permita a los colaboradores aplicar sus habilidades de manera sistemática.

### Punto de revisión de fase

Revisión al final de una fase en la que se toma una decisión de continuar a la siguiente fase, continuar con modificaciones o dar por concluido un programa o proyecto.

CRISP-DM define claramente actividades que determinan las transiciones entre fases.

Las conclusiones de este trabajo de investigación deben proveer un marco de trabajo que permita la evaluación de los procesos al final de cada fase, de igual manera debe permitir planificar lecciones aprendidas para optimizar el desempeño del equipo en siguientes etapas.

### Procesos de la dirección de proyectos

Serie sistemática de actividades dirigidas a producir un resultado final de forma tal que se actuará sobre una o más entradas para crear una o más salidas.

CRISP-DM está enfocada primordialmente a las actividades requeridas para la elaboración del producto final. Este modelo de procesos no contempla actividades que contribuyan a la gestión efectiva de los riesgos, la gestión de costos, aseguramiento de la calidad, incentivo a la gestión del conocimiento, planificación de comunicaciones efectivas, control de cambios, cronogramas y adquisiciones.

### Grupo de procesos de la dirección de proyectos

Agrupamiento lógico de las entradas, herramientas, técnicas y salidas relacionadas con la dirección de proyectos. Los grupos de procesos de la dirección de proyectos incluyen procesos de inicio, planificación, ejecución, monitoreo y control, y cierre.

CRISP-DM agrupa en fases las actividades que conforman el modelo de procesos. Considera actividades para cada grupo de procesos que describe el PMBOK.

Para fines de objetiva evaluación, este trabajo clasifica las mejores prácticas consideradas en el PMBOK por grupos de procesos y las compara con las actividades conocidas de CRISP-DM.



## Área de conocimiento de la dirección de proyectos

Área identificada de la dirección de proyectos definida por sus requisitos de conocimientos y que se describe en términos de sus procesos, prácticas, datos iniciales, resultados, herramientas y técnicas que los componen.

Los procesos se clasifican en la guía del PMBOK bajo las siguientes diez áreas de conocimiento:

- Gestión de la Integración del Proyecto
- Gestión del Alcance del Proyecto
- Gestión del Cronograma del Proyecto
- Gestión de los Costos del Proyecto
- Gestión de la Calidad del Proyecto
- Gestión de los Recursos del Proyecto
- Gestión de las Comunicaciones del Proyecto
- Gestión de los Riesgos del Proyecto
- Gestión de las Adquisiciones del Proyecto
- Gestión de los Interesados del Proyecto

CRISP-DM no segmenta las actividades por área de conocimientos. Está fuera del alcance de este trabajo clasificar las actividades por área de conocimientos.

Los líderes de proyectos deben poseer un conocimiento lo más detallado posible de cada una de estas áreas citadas, tanto en lo teórico como en lo práctico; conocer sus procesos, técnicas y herramientas.

### *Gestión de la Integración del Proyecto*

A través de esta área de conocimiento debe reflejarse el compromiso de los miembros del equipo durante la evolución del proyecto. Los miembros del equipo determinan como deben integrarse los planes y componentes. El enfoque del director del proyecto ha de ser la preparación de un ambiente colaborativo que permita la ejecución de las mejores decisiones tomadas por el equipo para el desarrollo del producto. El equipo debe poseer habilidades competitivas y experiencia de trabajo en grupos de trabajo o debe haber participado en su mayoría en proyectos similares previamente.

<b>Gestión de la Integración del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	Desarrollar el Acta de Constitución del Proyecto
Grupo de Procesos de Planificación	Desarrollar el Plan para la Dirección del Proyecto
Grupo de Procesos de Ejecución	Dirigir y Gestionar el Trabajo del Proyecto
	Gestionar el Conocimiento del Proyecto
Grupo de Procesos de Monitoreo y Control	Monitorear y Controlar el Trabajo del Proyecto

	Realizar el Control Integrado de Cambios
Grupo de Procesos de Cierre	Cerrar el Proyecto o Fase

*Tabla 1. Procesos Gestión de la Integración del Proyecto*

### *Gestión del Alcance del Proyecto*

En proyectos de este tipo con requerimientos que cambian con frecuencia, con altos riesgos e incertidumbre en su exitosa puesta en marcha, el Alcance regularmente no es determinado con claridad en fases iniciales del proyecto o cambia a medida que transcurre el tiempo. En metodologías ágiles normalmente no se determina y establece un criterio de aceptación definitivo antes de iniciar con el desarrollo del producto. En el enfoque adaptativo se debe redefinir los requerimientos en el inicio de cada iteración, revisar prototipos y generar versiones de cada liberación. El Alcance se redefine a lo largo del proyecto. La base fundamental del alcance y la definición de los requerimientos en proyectos adaptativos es el *backlog*.

<b>Gestión del Alcance del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión del Alcance
	Recopilar Requisitos
	Definir el Alcance
	Crear la EDT/WBS
Grupo de Procesos de Ejecución	
Grupo de Procesos de Monitoreo y Control	Validar el Alcance
	Controlar el Alcance
Grupo de Procesos de Cierre	

*Tabla 2. Procesos Gestión del Alcance del Proyecto*

### *Gestión del Cronograma del Proyecto*

Utilizar ciclos cortos para realizar el trabajo, revisar los resultados y aplicar cambios si es necesario. Estos ciclos han de proveer rápida retroalimentación sobre la idoneidad y efectividad de los entregables, lo cual manifiesta una programación iterativa de tareas en base a demandas y recomendaciones por parte de los usuarios.

<b>Gestión del Cronograma del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión del Cronograma
	Definir las Actividades

	Secuenciar las Actividades
	Estimar la Duración de las Actividades
	Desarrollar el Cronograma
Grupo de Procesos de Ejecución	
Grupo de Procesos de Monitoreo y Control	Controlar el Cronograma
Grupo de Procesos de Cierre	

*Tabla 3. Procesos Gestión del Cronograma del Proyecto*

### *Gestión de los Costos del Proyecto*

En proyectos con alto grado de incertidumbre en los cuales no es posible definir claramente el alcance, no podrán contar con estimaciones detalladas de costos. Estimaciones preliminares se pueden utilizar en fases iniciales para generar una rápida proyección a alto nivel de los costos del proyecto, los cuales pueden ir ajustándose a medidas en que van apareciendo los cambios.

En caso de contar con un presupuesto determinado, el alcance y la programación de actividades deben ajustarse a las limitaciones establecidas.

<b>Gestión de los Costos del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión de los Costos
	Estimar los Costos
	Determinar el Presupuesto
Grupo de Procesos de Ejecución	
Grupo de Procesos de Monitoreo y Control	Controlar los Costos
Grupo de Procesos de Cierre	

*Tabla 4. Procesos Gestión de los Costos del Proyecto*

### *Gestión de la Calidad del Proyecto*

Frecuentes tareas en cada iteración deben realizarse para asegurar la calidad de la ejecución de los procesos de gestión y de desarrollo del producto para que cumpla con las expectativas definidas previo a cada etapa iterativa.

Es recomendable que en los primeros entregables se incorporen la mayor parte de los elementos posibles para así descubrir posibles inconsistencias y problemas de calidad frente a las expectativas de los usuarios con el objetivo de afrontarlos en las tempranas iteraciones que representan un menor costo en aplicar las correcciones necesarias.

<b>Gestión de la Calidad del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión de la Calidad
Grupo de Procesos de Ejecución	Gestionar la Calidad
Grupo de Procesos de Monitoreo y Control	Controlar la Calidad
Grupo de Procesos de Cierre	

*Tabla 5. Procesos Gestión de la Calidad del Proyecto*

### *Gestión de los Recursos del Proyecto*

Los proyectos con alta variabilidad se benefician de las estructuras de equipos que maximizan su enfoque y colaboración.

La colaboración aumenta la productividad y facilita la resolución de problemas de forma innovadora. Los equipos colaborativos facilitan la rápida integración de personal de diversas áreas, tienen una mejor comunicación, se comparte el conocimiento adquirido y provee flexibilidad de asignación de trabajo, entre otras ventajas.

La colaboración en equipos es crítica en el éxito de los proyectos con alto grado de cambios, principalmente porque se pierde menos tiempo para que una función centralizada deba tomar decisiones de asignación de tareas.

<b>Gestión de los Recursos del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión de Recursos
	Estimar los Recursos de las Actividades
Grupo de Procesos de Ejecución	Adquirir Recursos
	Desarrollar el Equipo
	Dirigir al Equipo
Grupo de Procesos de Monitoreo y Control	Controlar los Recursos
Grupo de Procesos de Cierre	

*Tabla 6. Procesos Gestión de los Recursos del Proyecto*

### *Gestión de las Comunicaciones del Proyecto*

Los ambientes de proyectos sujetos a elementos de ambigüedad y cambios tienen la clara necesidad de comunicarse de manera más frecuente y de forma rápida.

Además, mantener los artefactos del producto de software y documentación del proyecto de forma transparente, notificación de estados de avances, y establecer

reuniones regulares de seguimiento promueven la comunicación y facilitan la gestión de los involucrados.

<b>Gestión de las Comunicaciones del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión de las Comunicaciones
Grupo de Procesos de Ejecución	Gestionar las Comunicaciones
Grupo de Procesos de Monitoreo y Control	Monitorear las Comunicaciones
Grupo de Procesos de Cierre	

*Tabla 7. Procesos Gestión de las Comunicaciones del Proyecto*

### *Gestión de los Riesgos del Proyecto*

Proyectos con alto grado de incertidumbre, normalmente cuentan con mayor existencia de riesgos. El riesgo ha de ser identificado, analizado y gestionado en cada iteración del proyecto.

Los requerimientos que representan un alto impacto no considerado en fases iniciales se conservan como documentos vivos que son actualizados regularmente, y el trabajo ha de ser re-priorizado a medida en que avanza el proyecto, basado en el entendimiento mejorado del nivel de riesgo que representan.

<b>Gestión de los Riesgos del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión de los Riesgos
	Identificar los Riesgos
	Realizar el Análisis Cualitativo de Riesgos
	Realizar el Análisis Cuantitativo de Riesgos
	Planificar la Respuesta a los Riesgos
Grupo de Procesos de Ejecución	Implementar la Respuesta a los Riesgos
Grupo de Procesos de Monitoreo y Control	Monitorear los Riesgos
Grupo de Procesos de Cierre	

*Tabla 8. Procesos Gestión de los Riesgos del Proyecto*

### *Gestión de las Adquisiciones del Proyecto*

En caso de necesitar extender el equipo con proveedores externos u obtener recursos no disponibles en la organización a través de otras entidades, es requerido llevar a cabo tareas de gestión para una efectiva colaboración a lo largo del proyecto.

Acuerdos para trabajar en modo adaptativo se pueden establecer con los proveedores con el objetivo de tener sus intervenciones en los momentos requeridos.

<b>Gestión de las Adquisiciones del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	
Grupo de Procesos de Planificación	Planificar la Gestión de las Adquisiciones
Grupo de Procesos de Ejecución	Efectuar las Adquisiciones
Grupo de Procesos de Monitoreo y Control	Controlar las Adquisiciones
Grupo de Procesos de Cierre	

*Tabla 9. Procesos Gestión de las Adquisiciones del Proyecto*

### *Gestión de los Interesados del Proyecto*

Proyectos con alto grado de cambios requieren de compromiso y participación con los involucrados. A menudo, el cliente, el usuario y desarrolladores intercambian informaciones de forma dinámica co-creativa que hace que los involucrados tengan más satisfacción por identificarse con el proyecto y ser parte de la solución. Interacciones regulares con los involucrados a lo largo del proyecto ayuda a mitigar los riesgos, fomenta la confianza y contribuye a que la mayor parte de los cambios se den en fases tempranas, lo cual representa una disminución en los costos y el incremento de la probabilidad de éxito del proyecto.

<b>Gestión de los Interesados del Proyecto</b>	
<b>Grupo de procesos</b>	<b>Proceso</b>
Grupo de Procesos de Inicio	Identificar a los Interesados
Grupo de Procesos de Planificación	Planificar el Involucramiento de los Interesados
Grupo de Procesos de Ejecución	Gestionar la Participación de los Interesados
Grupo de Procesos de Monitoreo y Control	Monitorear el Involucramiento de los Interesados
Grupo de Procesos de Cierre	

*Tabla 10. Procesos Gestión de los Interesados del Proyecto*

## 4.5 Comparación entre el Modelo de Procesos CRISP-DM y la Guía del PMBOK

Los fundamentos para la dirección de proyectos incluyen prácticas tradicionales comprobadas y ampliamente utilizadas, así como prácticas innovadoras emergentes para la profesión, (Project Management Institute, 2017a). Considerando que esta tarea consiste en la aplicación de conocimientos, habilidades, herramientas y técnicas a las actividades del proyecto para cumplir con los requisitos del mismo, se presenta en esta sección una evaluación de la aplicabilidad de los procesos requeridos para lograr el éxito en los proyectos relacionados al área de estudio.

A continuación, se evalúan los procesos descritos por el PMI en la guía del PMBOK frente a las etapas y actividades que conforman el modelo de procesos CRISP-DM.

PMBOK			CRISP-DM
Grupos de procesos	Procesos	Requerido	Etapas-Tarea
Inicio		✓	EN-Determinar Objetivos del Negocio
	Desarrollar el Acta de Constitución del Proyecto	✓	
	Identificar a los Interesados	✓	EN-Evaluar la situación
Planificación	Desarrollar el Plan para la Dirección del Proyecto	✓	
	Planificar la Gestión del Alcance	✓	
	Recopilar Requisitos	✓	EN-Evaluar la situación
	Definir el Alcance	✓	EN-Determinar Metas de Minería de Datos
	Crear la EDT/WBS		EN-Producir el Plan de Proyecto
	Planificar la Gestión del Cronograma	✓	EN-Producir el Plan de Proyecto
	Definir las Actividades	✓	EN-Producir el Plan de Proyecto
	Secuenciar las Actividades		EN-Producir el Plan de Proyecto
	Estimar la Duración de las Actividades	✓	EN-Producir el Plan de Proyecto
	Desarrollar el Cronograma	✓	EN-Producir el Plan de Proyecto
	Planificar la Gestión de los Costos	✓	
	Estimar los Costos	✓	
	Determinar el Presupuesto	✓	
	Planificar la Gestión de la Calidad	✓	
	Planificar la Gestión de Recursos	✓	
	Estimar los Recursos de las Actividades		
	Planificar la Gestión de las Comunicaciones	✓	
	Planificar la Gestión de los Riesgos	✓	
	Identificar los Riesgos	✓	EN-Evaluar la Situación
	Realizar el Análisis Cualitativo de Riesgos	✓	EN-Evaluar la Situación
	Realizar el Análisis Cuantitativo de Riesgos	✓	EN-Evaluar la Situación
	Planificar la Respuesta a los Riesgos	✓	EN-Evaluar la Situación
	Planificar la Gestión de las Adquisiciones	✓	
	Planificar el Involucramiento de los Interesados	✓	
			ED-Recoger Datos Iniciales
			ED-Describir los Datos
			ED-Explorar los Datos
			ED-Verificar la Calidad de los Datos
Ejecución	Dirigir y Gestionar el Trabajo del Proyecto	✓	
	Gestionar el Conocimiento del Proyecto	✓	
	Gestionar la Calidad	✓	
	Adquirir Recursos	✓	
	Desarrollar el Equipo	✓	

	Dirigir al Equipo	✓	
	Gestionar las Comunicaciones	✓	
	Implementar la Respuesta a los Riesgos	✓	
	Efectuar las Adquisiciones	✓	
	Gestionar la Participación de los Interesados	✓	
			PD-Seleccionar los Datos
			PD-Limpiar los Datos
			PD-Construir los Datos
			PD-Integrar los Datos
			PD-Formatear los Datos
			CM-Seleccionar la técnica de Modelado
			CM-Generar Diseños de Pruebas
			CM-Construir Modelo
			CM-Evaluar Modelo
			EP-Evaluar Resultados
<b>Monitoreo y Control</b>	Monitorear y Controlar el Trabajo del Proyecto	✓	
	Realizar el Control Integrado de Cambios	✓	
	Validar el Alcance	✓	
	Controlar el Alcance	✓	
	Controlar el Cronograma	✓	
	Controlar los Costos	✓	
	Controlar la Calidad	✓	
	Controlar los Recursos	✓	
	Monitorear las Comunicaciones		
	Monitorear los Riesgos	✓	
	Controlar las Adquisiciones	✓	
	Monitorear el Involucramiento de los Interesados	✓	
<b>Cierre</b>	Cerrar el Proyecto o Fase	✓	PP-Planificar Despliegue
			PP-Planificar Monitoreo y Mantenimiento
			PP-Elaborar el Reporte Final
			PP-Revisar Proyecto

EN: Entendimiento del negocio  
 ED: Entendimiento de los datos  
 PD: Preparación de los datos  
 CM: Creación de los modelos  
 EP: Evaluación y pruebas  
 PP: Despliegue o puesta en producción

**Tabla 11.** Comparación Procesos PMBOK y tareas del modelo de procesos CRISP-DM

En la **Tabla 11** previa se presenta cuales procesos de la guía de conocimientos del PMBOK son requeridos para la gestión efectiva de proyectos de Minería de Datos orientados a utilizar modelos de predicción de comportamientos. Esta comparación de muestra también la equivalencia de los procesos del PMBOK frente a las etapas y actividades que conforman el modelo CRISP-DM.

Se puede apreciar a partir de este análisis que 22 procesos considerados como buenas prácticas en dirección de proyectos no presentan actividades equivalentes en CRISP-DM. Mientras que 17 actividades de CRISP-DM no forman parte de las recomendaciones del PMBOK; esto por ser propias a tareas de Minería de Datos.



## 4.6 Metodologías ágiles

El 17 de febrero de 2001 diecisiete críticos de los modelos de desarrollo de software basados en procesos, encabezados por Kent Beck, se reunieron en Snowbird, Utah para discutir sobre técnicas y procesos para la industria del desarrollo de software. A partir de esta reunión nace la filosofía ágil en la dirección de proyectos como alternativa de las metodologías formales que surgían en ese entonces (CMMI, SPICE, entre otras) a las que consideraban excesivamente rígidas por su carácter normativo y fuerte dependencia de planificaciones detalladas previas al desarrollo.

Los integrantes de la reunión resumieron los principios sobre los que se basan los métodos alternativos en lo que se denomina como Manifiesto Ágil.

En el Manifiesto Ágil se ofrece una descripción general de los fundamentos de las metodologías ágiles, (Crisp y McKenna, 2016):

- El fundamento de las metodologías ágiles consiste en **dar prioridad a las personas e iteraciones en lugar de los procesos y herramientas**. Las metodologías resultan ser flexibles y orientadas a la realidad de trabajo que enfrentan los equipos en las organizaciones.
- Otro aspecto que se destaca en las herramientas ágiles es el enfoque a la: **Colaboración del cliente sobre la Negociación vía Contrato**. Esta visión promueve la confianza entre las partes involucradas para alcanzar los objetivos del proyecto.
- **Responden a los Cambios en lugar de Seguir un Plan**. Es otra perspectiva de las metodologías ágiles. En metodologías basadas en ciclo de vida en Cascada, la mayor parte de la planificación se realiza en etapas tempranas del proyecto, en cambio, en un enfoque ágil los planes son desarrollados durante el desarrollo del proyecto de forma iterativa, lo cual permite conocer efectivamente las características más relevantes para el desarrollo del producto. En metodologías prescriptivas es frecuente detectar errores en el producto en fases finales, lo cual hace más costoso reparar los daños causados por un error en las fases iniciales.
- Las metodologías ágiles prestan más importancia al **Software funcionando en lugar de una documentación extensiva**. Las metodologías prescriptivas promueven la formalidad en los procesos, lo cual da como resultado amplia documentación.

Los siguientes valores sirven de guía a la hora de aplicar la filosofía ágil de trabajo en los proyectos de desarrollo de software, (Crisp y McKenna, 2016):

### La prioridad es satisfacer al cliente de manera temprana y entregar continuamente software de valor

La meta es no solo producir algo que el cliente quiere, sino algo que él no pueda esperar a utilizarlo.

En enfoques predictivos el cliente debe esperar a que concluya una serie de etapas burocráticas para obtener lo que ha solicitado. Proyectos de minería de datos para identificar patrones de comportamiento y así captar nuevos clientes resultan ser volátiles, y lo que la organización pudo necesitar en la fase de concepción de la idea puede que cambie durante el desarrollo del producto.

### Aceptación de cambios incluso en fases finales del desarrollo del producto

En las metodologías ágiles se enfoca en la adaptación de los cambios, en lugar de planearlo todo. El cliente no puede conocer todos los requisitos en fases iniciales, sino que las concepciones iniciales cambian constantemente durante el desarrollo. El equipo debe trabajar de forma tal que pueda esperar y aceptar los cambios.

CRISP-DM no considera un enfoque adaptativo, lo cual puede no ser lo suficientemente efectivo para alcanzar las metas de la organización que es lo primordial: desarrollar modelos para ganar nuevos clientes.

### Entregar software en funcionamiento con frecuencia, desde un par de semanas o un par de meses, con preferencia a escalas de tiempo más cortas

Es una costumbre no presentar un producto hasta que no haya sido finalizado. En desarrollo de software muchas veces no es conveniente, debido a que esa pequeña funcionalidad puede aprovecharse para generar valor al proyecto.

Los ciclos de entregas son denominados *Sprints*. Estos deben ser programados lo más cortos posibles para tener retroalimentación de los involucrados y determinar el avance del proyecto.

## El personal de Negocios y desarrolladores deben trabajar juntos todos los días del proyecto

La idea detrás de un proyecto de software es satisfacer el Negocio. Cuando se mantiene los equipos trabajando juntos, hace que trabajen enfocados en las metas de la organización.

Aunque no sea posible trabajar todos los días juntos, no obstante, el contacto debe ser lo más frecuente posible.

CRISP-DM no presenta principios o tareas que motiven el trabajo en equipo. La ejecución de sus actividades puede ser más efectiva con una visión de trabajo en equipo.

## Construir proyectos alrededor de personas motivadas. Brindarles el entorno y apoyo que necesitan y confiar en que hagan el trabajo bien

El equipo de trabajo es lo más importante. Se debe contar con un equipo de personas que complemente sus habilidades para completar satisfactoriamente el trabajo asignado.

En un enfoque predictivo el director establecería el qué y cómo hacer el trabajo, mientras que en un entorno ágil el *Product Owner* daría las pautas de las necesidades y el equipo establecería la mejor manera de hacer el trabajo.

## El método más eficiente y eficaz para transmitir información a un equipo de desarrollo y dentro de él es la conversación cara-a-cara

La comunicación cara-a-cara es la forma más efectiva debido a que todas las facetas de comunicación están presentes. Expresiones faciales, tono de voz, lenguaje corporal y palabras habladas.

CRISP-DM no contempla planes de gestión de comunicaciones, tampoco actividades que fomenten la integración del equipo y utilicen el método de comunicación cara-a-cara.

## Software funcional es la principal medida de progreso

El manifiesto ágil establece que el software funcional debe ser el enfoque del equipo, en lugar de documentaciones, diseño o especificaciones.

El software no está completo cuando ya esté terminado el trabajo previsto, sino cuando “funcione”, sea validado y demostrado que logra los objetivos.

Los modelos de minería de datos han de cumplir con este principio, de lo contrario no representaría valor para la organización.

## Los procesos ágiles promueven el desarrollo sostenible. Los patrocinadores, desarrolladores y usuarios deberían poder mantener un ritmo constante indefinidamente

Las metodologías ágiles promueven un ritmo constante de trabajo iterativo independientemente de fechas de compromisos que deban cambiar el modelo exitoso de trabajo en equipo.

## Atención continua a la excelencia técnica y al buen diseño aumenta la agilidad

El uso de estándares, prácticas reconocidas, algoritmos altamente probados en la industria, procurar la legibilidad del código de software, diseños que permitan fácilmente ser actualizados contribuye a que el equipo no retrabaje o pierda tiempo aplicando correcciones y mejoras en las iteraciones.

## La simplicidad, el arte de maximizar la cantidad de trabajo no realizado. Es esencial

Un equipo ágil debe construir las funcionalidades lo más simple posible. Esto evita la pérdida de tiempo al momento de revisar y aplicar cambios a funcionalidades previamente construidas.

## Las mejores arquitecturas, requisitos y diseños surgen de equipos auto-organizados

Permitir la libertad a los equipos para que trabajen en la manera en que se sienta más cómodos. Esto repercutirá positivamente en los resultados de los proyectos y la organización contará con un personal motivado y confiado en el valor que aporta su colaboración.

## A intervalos regulares, el equipo reflexiona sobre cómo ser más efectivo, luego optimiza y ajusta

Las metodologías ágiles tienen que ver con la mejora continua. Un equipo ágil inspecciona todo lo que hace y rutinariamente aplica cambios en un esfuerzo de mejorar su forma de trabajo. Evaluar la forma de trabajo una vez concluido el proyecto es beneficioso, pero solo para futuros proyectos. Es vital que exista un monitoreo constante que evalúe la manera en que trabajamos.

CRISP-DM no estipula dentro de sus fases, planes de seguimiento y evaluación de las operaciones del equipo de trabajo con el objetivo de mantener una optimización en su forma de trabajo.

## 4.7 Marco de trabajo ágil Scrum

El ciclo de vida de un proyecto es la serie de fases que atraviesa un proyecto desde su inicio hasta su conclusión. Proporciona el marco de referencia básico para dirigir el proyecto. Las fases pueden ser secuenciales, iterativas o superpuestas, (Project Management Institute, 2017a). El marco de trabajo Scrum presenta un ciclo de vida con fases superpuestas en un entorno adaptativo.

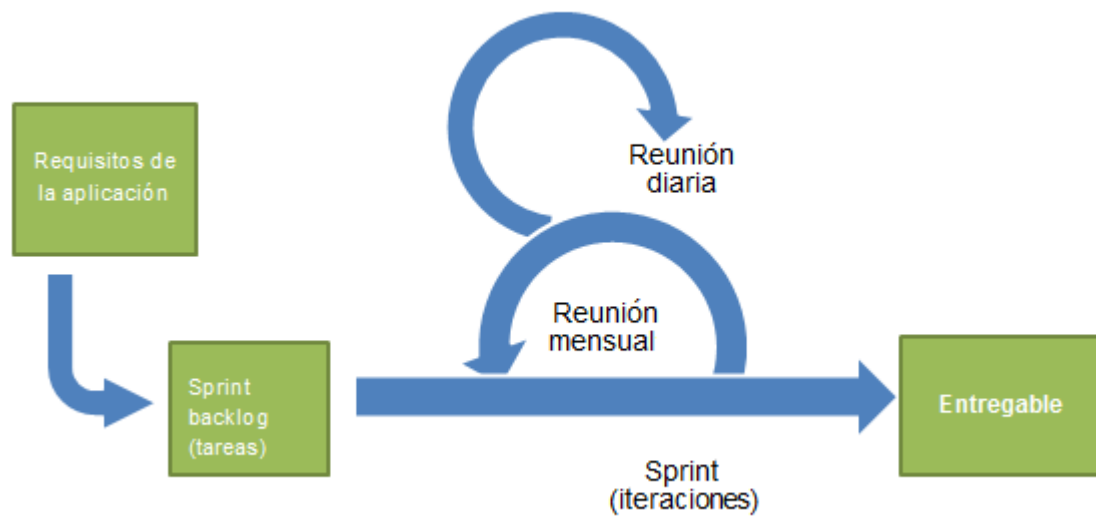
SCRUM Es un marco de trabajo a través del cual las personas pueden abordar problemas complejos adaptativos, a la vez que se entregan productos de forma eficiente y creativa con el máximo valor, (Schwaber y Sutherland, 2016).

Uno de los aspectos de Scrum es su simplicidad. A pesar de la ligereza de esta metodología, cada uno de sus elementos responde a la problemática que se presenta en las metodologías tradicionales. Scrum es mucho más sobre personas, comportamientos y cultura. La mentalidad y principios relacionados a las metodologías ágiles indican un implícito entorno sobre el cual esta metodología prosperará.

La implementación de Scrum puede parecer simple por sus pocas especificaciones y las pocas restricciones al compararse con enfoques predictivos. El mayor reto está en adoptar una postura de compromiso, de trabajo en equipo y promover una cultura organizacional que siga los principios propios de las metodologías ágiles.

Scrum es un marco de trabajo en la que se aplican de manera regular un conjunto de buenas prácticas para **trabajar colaborativamente, en equipo**, y obtener el mejor resultado posible de un proyecto. Estas prácticas se apoyan unas a otras y su selección tiene origen en un estudio de la manera de trabajar de equipos altamente productivos. Se pueden emplear un conjunto de diferentes procesos y técnicas para el fin de adaptar a la organización y desarrollar los proyectos.

Los equipos Scrum, roles, eventos, artefactos y reglas asociadas componen este marco de trabajo. Cada componente sirve para un propósito específico y es esencial para el éxito en su implementación como marco de trabajo.



*Ilustración 5. Marco de Trabajo Scrum*

## Usos de Scrum

Scrum inicialmente fue desarrollado para gestionar y desarrollar productos de software. A partir de la década de 1990s, Scrum se ha utilizado extensivamente en todo el mundo para, (Schwaber y Sutherland, 2016):

1. Investigar e identificar mercados viables, tecnologías, y capacidades;
2. Desarrollo de productos y mejoras;
3. Lanzamientos de productos y mejoras, diariamente tantas veces como sea posible;
4. Desarrollo y mantenimiento en la Nube (online, seguridad, por-demanda) y otros entornos operacionales de desarrollo para el uso de producto; y,
5. Mantenimiento y renovación de productos.

Hoy en día ha sido considerado como base en el emprendimiento de proyectos de Inteligencia de Negocios y Minería de datos, (Muntean y Surcel, 2013).

## Teoría de Scrum

Scrum se basa en la teoría de control de procesos empírica o empirismo. El empirismo asegura que el conocimiento procede de la experiencia y en poder tomar decisiones basándose en lo conocido. Scrum emplea un enfoque iterativo e incremental para optimizar la predictibilidad y el control del riesgo. Tres pilares soportan toda la implementación del control de procesos empírico: transparencia, inspección y adaptación.

## El Equipo Scrum

El equipo Scrum (*Scrum Team*) se conforma de un Propietario del Producto (*Product Owner*), *equipo de desarrollo (Development Team)* y un *Scrum Master*.

### *Propietario del Producto (Product Owner)*

El Propietario del Producto es el representante del negocio, quien conoce las operaciones de la organización. Es quien gestiona el listado de requerimientos. Debe aclarar los objetivos del proyecto al equipo Scrum.

### *El Equipo de Desarrollo (Development Team)*

Representa el equipo de personas que realizan el trabajo de entregar un incremento de producto que potencialmente se pueda desplegar en ambiente productivo al final de cada *Sprint*.

Los Equipos de Desarrollo tienen las características siguientes:

- Auto-organizados
- Multifuncionales: Cuentan con todas las habilidades necesarias para crear un incremento de producto.
- No reconoce títulos para los miembros del equipo.
- No reconoce sub-equipos.
- Los miembros individuales del Equipo de Desarrollo pueden tener habilidades especializadas y áreas en las que estén más enfocados, sin embargo, la responsabilidad recae en el Equipo de Desarrollo como conjunto.

### *El Scrum Master (Scrum Master)*

De cara al Propietario del Producto, el Scrum Master debe asegurar que los objetivos y el alcance del producto sean entendidos por todos en el Equipo. Debe proveer



técnicas para gestionar los requerimientos del proyecto. Debe planificar el trabajo en un ambiente adaptativo por los cambios. Entender y aplicar la agilidad. Facilitar los eventos de Scrum según se requiera o necesite.

Con relación al Equipo de Desarrollo, debe guiarlos a ser auto-organizado, proveer las herramientas para crear productos de alto valor, eliminar impedimentos para el progreso del equipo. El Scrum Master debe facilitar los eventos Scrum según se requiera o necesite.

Respecto la Organización, el Scrum Master tiene la responsabilidad de liderar y guiar la Organización en la adopción de Scrum como marco de trabajo. Debe planificar las implementaciones de Scrum y motivar cambios que incrementen la productividad del Equipo Scrum.

### *Gerente del Producto (Product Manager)*

En los últimos tiempos ha sido incorporada un nuevo rol en el desarrollo de los proyectos. Se trata del Gerente del Producto (*Product Manager*). Representa el rol con la responsabilidad ejecutiva sobre los resultados del proyecto.

## Eventos Scrum

### *El Sprint*

Es el periodo de tiempo de un mes o menos durante el cual se crea un incremento de producto utilizable y potencialmente desplegable.

### *Scrum Diario (Daily Scrum)*

El Scrum Diario es una reunión de aproximadamente 15 minutos que se realiza cada día del Sprint para el Equipo de Desarrollo. En esta reunión se planifica el trabajo de las siguientes 24 horas. Representa un evento importante que motiva la colaboración y desempeño efectivo del equipo inspeccionando el trabajo avanzado desde el último Scrum Diario y proyectando el trabajo del Sprint a realizar a continuación

Todos de pie, cada integrante del equipo va respondiendo las siguientes preguntas:

- ¿Qué he completado desde el último *standup*?
- ¿Qué completaré a partir de este *standup* y el próximo?
- ¿Qué impedimentos he tenido?

### *Revisión del Sprint (Sprint Review)*

Las Revisiones de los Sprint son inspecciones de los incrementos que se realizan al final de cada Sprint. Se evalúa lo que se ha hecho durante el previo Sprint y se determina las acciones que pueden llevarse a cabo para mejorar.

- El Equipo de Desarrollo explica que estuvo bien durante el Sprint, qué problemas surgieron y como fueron resueltos.
- El Equipo de Desarrollo presenta el trabajo realizado.
- El Propietario del Producto habla sobre la Pila de Producto (*Product Backlog*) en su estado actual.
- El grupo completo colabora acerca de qué hacer a continuación.
- Revisión de la cronología, presupuesto, capacidades potenciales y mercado para la próxima entrega prevista del producto.

### *Retrospectiva del Sprint*

Es una oportunidad para el Equipo Scrum de inspeccionarse a sí mismo y de crear un plan de mejoras para el próximo Sprint.

## Artefactos Scrum

### *Pila del Producto (Product Backlog)*

Lista ordenada de requerimientos solicitados y administrados por el Propietario del Producto. Este documento puede cambiar a medida en que avanza el desarrollo del proyecto incluyendo nuevas ideas o ajustando las necesidades planteadas en un inicio. Los elementos de esta lista tienen como atributos la descripción del requerimiento, el orden de priorización, la estimación y el valor.

Los cambios en los requisitos de negocio, las condiciones del mercado o la tecnología pueden provocar cambios en la Pila del Producto.

### *Pila del Sprint (Sprint Backlog)*

Es el listado de los elementos de la Pila del Producto seleccionados para el Sprint. Consiste en el conjunto de funcionalidades que serán parte del incremento del producto al momento de haber concluido el Sprint.

## 4.8 Comparación entre el Modelo de Procesos CRISP-DM y la metodología ágil Scrum

A continuación, se evalúa las actividades asociadas a la metodología ágil Scrum regularmente utilizadas para abordar proyectos de Minería de Datos, frente a las etapas y tareas que conforman el modelo de procesos CRISP-DM.

Se presenta la comparación de las actividades que guardan relación con el ciclo de vida adaptativo de Scrum.

Scrum		CRISP-DM
Etapas	Actividad	Etapas-Tarea
Planificación	Exposición de prioridades	EN-Determinar Metas de Minería de Datos
		ED-Recoger Datos Iniciales
		ED-Describir los Datos
		ED-Explorar los Datos
		ED-Verificar la Calidad de los Datos
	Planificación del Sprint	
	Resolución de dudas	
Sprint	Definir objetivos del Sprint	
	Ciclo de desarrollo básico de Scrum, de duración máxima de 30 días en el que se desarrolla un incremento del producto	PD-Seleccionar los Datos
		PD-Limpiar los Datos
		PD-Construir los Datos
		PD-Integrar los Datos
		PD-Formatear los Datos
		CM-Seleccionar Técnica de Modelado
		CM-Generar Diseños de Pruebas
		CM-Construir Modelo
		CM-Evaluar Modelos
	Reunión diaria	
Incremento	Revisión del trabajo; resolución de trabas	EP-Evaluar los Resultados
		EP-Revisar el Proceso
	Presentación del incremento, sugerencias, anuncio próximo sprint	EP-Determinar los Próximos Pasos
	Retrospectiva	

EN: Entendimiento del negocio  
 ED: Entendimiento de los datos  
 PD: Preparación de los datos  
 CM: Creación de los modelos  
 EP: Evaluación y pruebas  
 PP: Despliegue o puesta en producción

**Tabla 12.** Comparación Actividades Scrum y tareas del modelo de procesos CRISP-DM

El análisis presentado en la tabla anterior revela la comparación entre las actividades de Scrum y las tareas de CRISP-DM que no guardan relación con los procesos de la guía del PMBOK.

Podemos interpretar a partir de estos datos que cinco de las actividades de CRISP-DM describen un enfoque de planificación, once corresponden a una fase de desarrollo del producto, y una a la revisión del trabajo realizado.

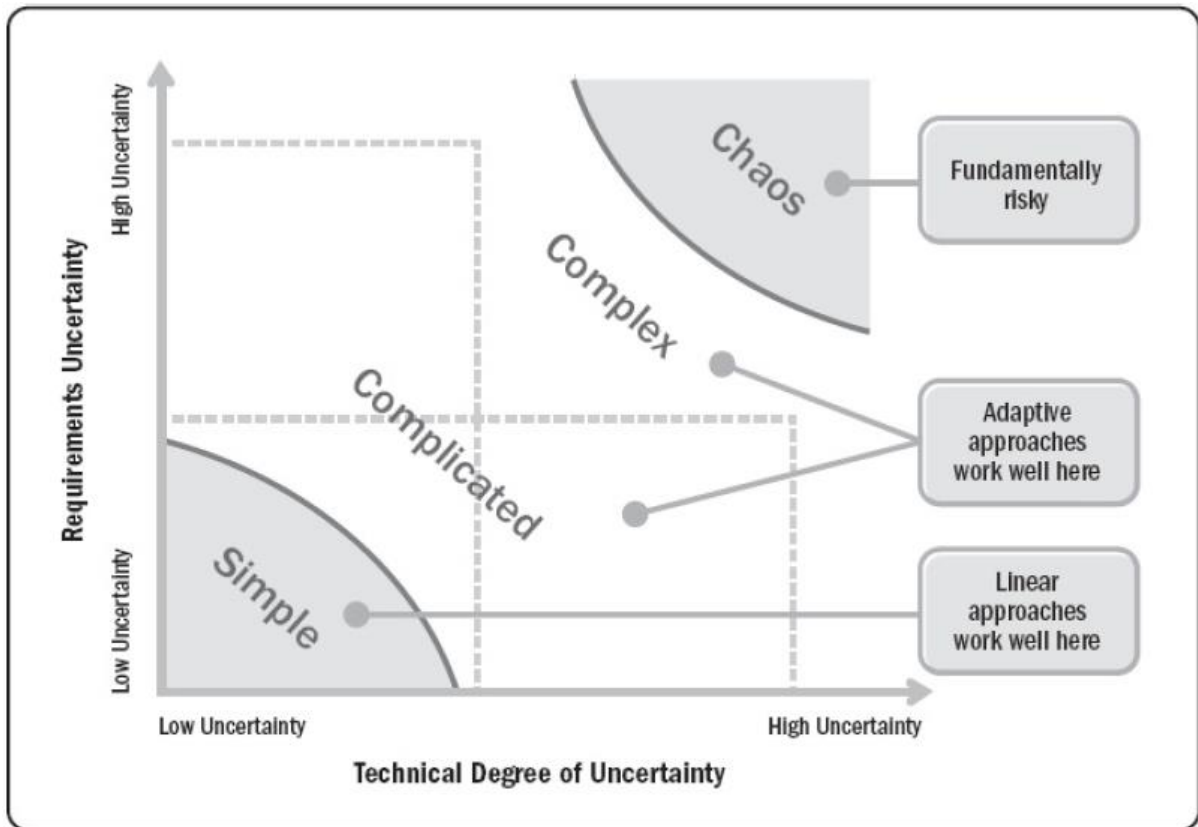
Las actividades de Scrum que no guardan equivalencia con el modelo de procesos CRISP-DM pueden considerarse como apropiadas para la gestión de los proyectos bajo un entorno ágil.

## 4.9 Propuesta de marco de trabajo para proyectos de Minería de Datos orientados a captación de clientes

Una vez analizados los modelos de trabajos más importantes en la industria del desarrollo de software, las recomendaciones de gestión de proyectos descritas en la literatura, las prácticas de desarrollo de modelos predictivos de Minería de Datos, nuestra experiencia laboral y académica, podemos identificar debilidades en los enfoques de trabajos utilizados en la actualidad para proyectos de captaciones de clientes vía reconocimientos de patrones de comportamientos.

Según (Project Management Institute, 2017b), es recomendable guiarse por enfoques iterativos, incrementales y ágiles en proyectos que involucran herramientas, técnicas o dominios de aplicación innovadores. Este enfoque aplica de igual modo para proyectos que requieren de investigación y desarrollo, que cuenten con alta tasa de cambios, tengan requisitos ambiguos o que la meta del producto final sea difícil de describir. Proyectos para captar clientes, utilizando como base datos históricos para detectar patrones de comportamientos cumplen con estas condiciones.

Este trabajo recomienda una adaptación del marco de trabajo CRISP-DM considerando las prácticas descritas en el PMBOK bajo un enfoque ágil como lo es Scrum. Un modelo híbrido de trabajo que combine la estimación y gestión de seguimiento global al progreso del proyecto como instancia predictiva, teniendo en cuenta la flexibilidad e iteraciones que proporcionan las metodologías ágiles para el ciclo de vida del desarrollo del producto.



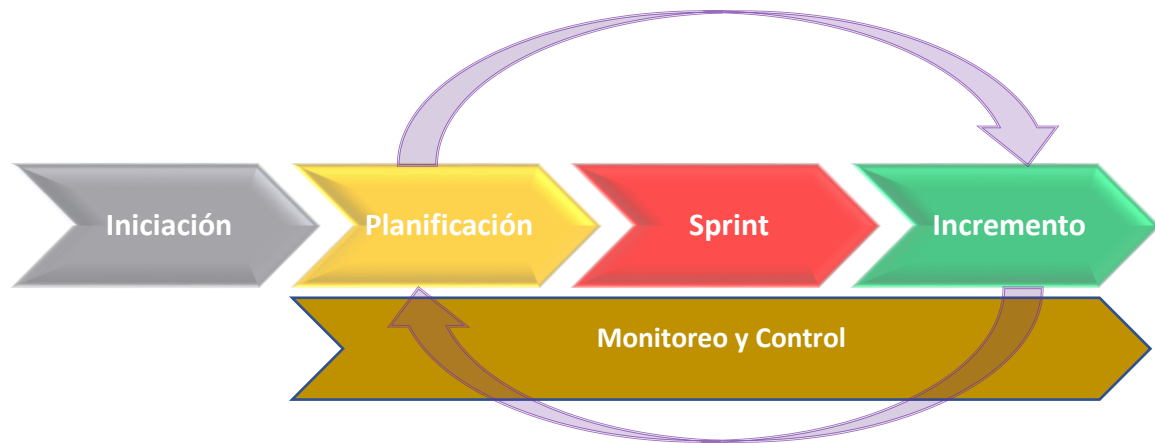
*Ilustración 6. Modelo de Incertidumbre y complejidad inspirado por el Modelo de Complejidad Stacey, (Project Management Institute, 2017b)*

La adaptación es necesaria porque se requiere un enfoque predictivo para definir las estrategias de gestión sobre los componentes que no representan alta variabilidad o representan indicadores importantes en la evaluación del proyecto en la Organización.

Se recomienda un enfoque iterativo y ágil que promueva el compromiso de los miembros del equipo como expertos locales en la gestión de la integración. En esta propuesta, el director del proyecto junto a los demás miembros del equipo determina cómo han de integrarse planes y componentes de manera empírica en la mayoría de casos.

Con este marco de trabajo proponemos invertir menos tiempo tratando de definir y acordar el alcance total en etapas tempranas del proyecto con el objetivo de invertir más tiempo en el descubrimiento de conocimiento y perfeccionamiento constante, ya que proyectos como estos suelen revelar brechas entre las necesidades reales de negocios y los requisitos que son señalados en un principio.

Se propone una fase de Iniciación con enfoque predictivo en la que se define el alcance al nivel más detallado que sea posible con la información disponible, y una fase de Planificación dentro de un ciclo iterativo y adaptativo en el cual dicho alcance es redefinido. Una fase transversal de Monitoreo y Control se incorpora con la finalidad de poner en práctica medidas que brinden seguridad al proyecto y permita mantener informada a la alta gerencia sobre su avance.



*Ilustración 7. Propuesta de Marco de Trabajo*

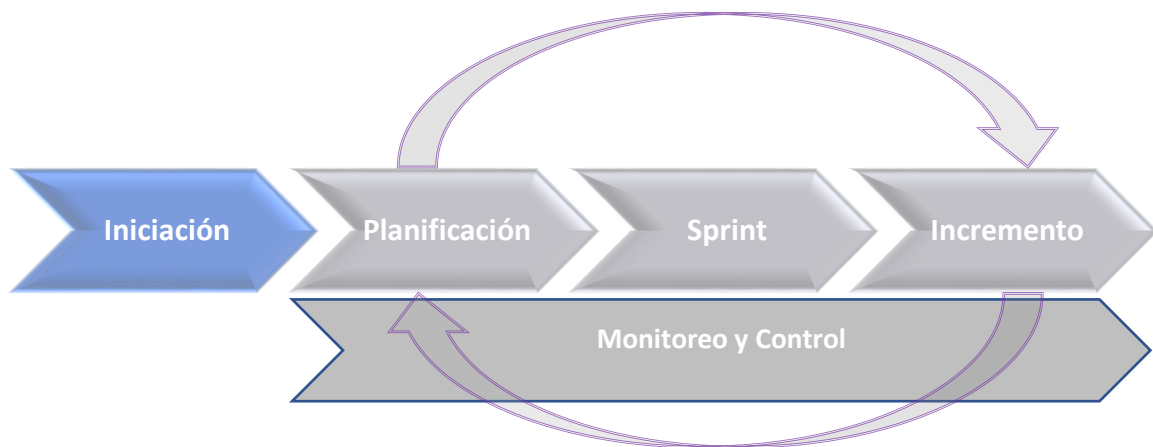
Iniciación		Planificación	Sprint	Incremento
EN-Determinar Objetivos del Negocio		ED-Recoger Datos Iniciales	PD-Seleccionar los Datos	EP-Determinar los Próximos Pasos
EN-Evaluar la situación		ED-Describir los Datos	PD-Limpiar los Datos	PP-Planificar Despliegue
EN-Determinar Metas de Minería de Datos			PD-Construir los Datos	PP-Planificar Monitoreo y Mantenimiento
Desarrollar el Acta de Constitución del Proyecto		ED-Explorar los Datos	PD-Integrar los Datos	PP-Elaborar el Reporte Final
Planificar la Gestión de los Costos		ED-Verificar la Calidad de los Datos	PD-Formatear los Datos	PP-Revisar Proyecto
Estimar los Costos		Planificación del Sprint	CM-Seleccionar Técnica de Modelado	Retrospectiva
Determinar el Presupuesto		Resolución de dudas	CM-Generar Diseños de Pruebas	Validar el Alcance
Planificar la Gestión de la Calidad		Planificar la Gestión del Cronograma	CM-Construir Modelo	Controlar el Alcance
Planificar la Gestión de Recursos		Desarrollar el Cronograma	Reunión diaria	
Planificar la Gestión de las Comunicaciones		Definir las Actividades	CM-Evaluar Modelos	
Planificar la Gestión de los Riesgos		Estimar la Duración de las Actividades	EP-Evaluar los Resultados	
Planificar la Gestión de las Adquisiciones		Identificar los Riesgos	EP-Revisar el Proceso	
Planificar el Involucramiento de los Interesados		Realizar el Análisis Cualitativo de Riesgos Planificar la Respuesta a los Riesgos		
Monitoreo y Control	Dirigir y Gestionar el Trabajo del Proyecto			
	Gestionar el Conocimiento del Proyecto			
	Gestionar la Calidad			
	Adquirir Recursos			
	Desarrollar el Equipo			
	Dirigir al Equipo			
	Gestionar las Comunicaciones			
	Implementar la Respuesta a los Riesgos			
	Efectuar las Adquisiciones			
	Gestionar la Participación de los Interesados			
	Monitorear y Controlar el Trabajo del Proyecto			
	Realizar el Control Integrado de Cambios			
	Controlar el Cronograma			
	Controlar los Costos			
	Controlar la Calidad			
	Controlar los Recursos			
	Monitorear los Riesgos			
	Controlar las Adquisiciones			
	Monitorear el Involucramiento de los Interesados			

EN: Entendimiento del negocio  
ED: Entendimiento de los datos  
PD: Preparación de los datos  
CM: Creación de los modelos  
EP: Evaluación y pruebas  
PP: Despliegue o puesta en producción

**Tabla 13.** Modelo de Trabajo Metodología Propuesta

Sobre la base de las buenas prácticas descritas en esta prestigiosa guía de Dirección de Proyectos del PMI, tomando en cuenta el enfoque de las metodologías ágiles y las etapas ya conocidas del tradicional esquema de trabajo utilizado: CRIPS-DM, se presentan recomendaciones a las organizaciones respecto a las pautas para la dirección de proyectos relacionados a minería de datos que contribuyan a la captación de nuevos clientes. La siguiente propuesta representa una metodología, guía para la definición de políticas, procedimientos, reglas, uso de herramientas y técnicas, y fases del ciclo de vida necesarios para la práctica de la dirección de estos tipos de proyectos.

## Fase de Iniciación: Enfoque predictivo



*Ilustración 8. Propuesta de Marco de Trabajo -Iniciación-*

La fase de Iniciación representa el enfoque predictivo que se recomienda para la gestión de proyectos de Minería de Datos. Incorpora procesos que establece la visión del proyecto y recomendaciones del PMBOK del grupo de procesos de Planificación, los cuales encajan en la definición del marco de trabajo a realizar.

### *EN-Determinar los objetivos de Negocio*

Esta tarea, propia del marco de procesos CRISP-DM no guarda correlación con los procesos descritos en el PMBOK, tampoco respecto a las actividades descritas en la metodología Scrum. Presenta el enfoque de determinar las metas necesarias para alcanzar el éxito del proyecto.

Esta tarea puede considerarse en metodologías tradicionales como parte de la responsabilidad del equipo conocedor de las operaciones de la empresa para documentar el Documento de Negocio, el cual ha de servir de entrada al proceso Desarrollar Acta de Constitución del Proyecto.

### *Desarrollar el Acta de Constitución del Proyecto*

Este proceso, propio del PMBOK se incluye en la propuesta para dar formalidad al desarrollo del proyecto. Debe considerarse este proceso con el objetivo de conferir



al director del proyecto la responsabilidad de gestionar los recursos disponibles que provee la organización para lograr el éxito en el proyecto.

Se ofrece el documento **Error! Reference source not found.** como guía de redacción para el Acta de Constitución del proyecto.

#### *EN-Evaluar la situación*

Como parte de la fase de Iniciación, debe considerarse esta tarea que es parte de CRISP-DM. Se recomienda complementar esta actividad con el proceso de Identificar a los Interesados descrito en el PMBOK.

En adición a llevar a cabo tareas para identificar los recursos necesarios para el proyecto y explorar las asunciones de los involucrados, se recomienda a partir de las buenas prácticas del PMBOK, determinar cuáles son sus intereses, grado de participación en el proyecto, interdependencias entre ellos, influencia de uno sobre otros e impactos que pudiera incidir en el desarrollo del proyecto.

Esta tarea en CRISP-DM considera también la recopilación de requisitos para el proyecto a desarrollar. Como parte de esta guía, debe realizarse un levantamiento de información sobre los requisitos iniciales para fines de elaborar un plan general que irá ajustándose a medida en que ocurren las iteraciones.

#### *EN-Determinar Metas de Minería de Datos*

Esta tarea en CRISP-DM establece las metas técnicas sobre lo requerido por la Organización. Es una tarea que se mantiene en esta propuesta con el objetivo de complementar la recopilación de requisitos, expectativas de los involucrados y servir de referencia al equipo de desarrollo.

El proceso Definir el Alcance del PMBOK se considera parte de esta tarea. No existirá una descripción detallada del alcance del proyecto, sino una referencia para fines de seguimiento y control con el objetivo de proveer a la alta gerencia durante el desarrollo del proyecto el estatus y una proyección lo más cercana posible a la realidad respecto a los avances alcanzados. Debemos tomar en cuenta que el alcance real será determinado a medida en que se construyan los modelos al final de los Sprints.

Esta tarea es parte también de la fase de Planificación ya que en cada iteración puede surgir diversos objetivos técnicos.

### *Planificar la Gestión de los Costos*

La gestión de los costos no está considerada en CRISP-DM. Scrum no considera actividades de dirección en sus fases iterativas. Es por tanto que se toma la referencia del PMBOK para dar seguimiento a tan importante indicador de éxito en los proyectos.

Si bien es cierto que en ciclos de vida adaptativos y proyectos de Minería de Datos en los cuales no existe una planificación detallada en los inicios y se caracterizan por su volatilidad, es importante mantener una postura crítica sobre los costos en los cuales incurren estos proyectos durante su ejecución. Claramente habrá márgenes de error en los estados y proyecciones respecto a estimaciones iniciales, sin embargo, es posible dar seguimiento en base a los incrementos que van ocurriendo luego de cada Sprint.

Debe describirse la manera en que se planificarán, estructurarán y controlarán los costos del proyecto. Este seguimiento puede repercutir en la toma de decisión para continuar o detener el proyecto antes de que sea demasiado tarde.

Herramientas:

- Juicio de expertos
- Análisis de datos
- Reuniones

### *Estimar los Costos*

Este es otro proceso no considerado en los modelos de procesos estudiados en este trabajo de investigación. Se ha incluido en esta propuesta con el objetivo de desarrollar una aproximación de los recursos monetarios necesarios para completar el trabajo del proyecto.

Se recomienda establecer las estimaciones con unidad *horas/hombre*, ya que son las unidades que se designan a las actividades en la Pila del Sprint.

En otro orden, las estimaciones deben considerar los costos asociados por adquisición o renta de recursos de infraestructura tecnológica que se requieran para el proyecto.

Este proceso se incluye en la fase de Iniciación para proveer a la alta gerencia una estimación general sobre el costo del proyecto.

Herramientas:

- Juicio de expertos: Evaluar proyectos anteriores
- Estimación análoga: Comparar con proyectos anteriores
- Estimación paramétrica: DMCoMo y análisis de varianza (ANOVA)

$$y = a_0 + \sum_{i=1}^n a_i x_i + e_i,$$

### *Determinar el Presupuesto*

Una tarea importante a incluir en el modelo que se propone en esta guía de trabajo es determinar el presupuesto en base a la estimación del esfuerzo y recursos a utilizar durante el proyecto. Es una aproximación calculada en base a las estimaciones realizadas. El presupuesto es un insumo primordial al momento de controlar los costos en la evolución del proyecto.

Debe documentarse cuales son los fondos autorizados para el proyecto, incluyendo reservas para contingencias.

Herramientas

- Juicio de expertos
- Análisis de datos

### *Planificar la Gestión de la Calidad*

En fases iniciales podemos planificar la manera en que controlaremos la calidad del proyecto y entregables.

Identificar los principales atributos de conformidad o no conformidad respecto a los criterios de aceptación del Propietario del Producto y/o Gerente del Producto.

El Director de Proyecto debe establecer métricas que han de ser evaluadas durante el desarrollo del proyecto y presentadas a la alta gerencia y Gerente del Producto.

Herramientas:

- Juicio de expertos
- Recopilación de datos: Tormenta de ideas
- Recopilación de datos: Entrevistas
- Análisis de datos: Análisis costo-beneficio
- Análisis de datos: Costo de la calidad
- Reuniones

### *Planificar la Gestión de Recursos*

Aunque las fases principales del proyecto han de llevarse a cabo de modo iterativo y respondiendo a las necesidades que surjan en los Sprints, en la etapa de Iniciación recomendamos establecer las condiciones de disponibilidad de los recursos que con alta probabilidad necesitará el proyecto.

El Director de Proyecto ha de ser consiente del entorno de trabajo y su influencia sobre los recursos. El entorno de trabajo, ubicación geográfica del equipo, condiciones del mercado, las políticas y cultura laboral de la organización pueden incidir en los recursos.

Herramientas:

- Juicio de expertos
- Reuniones
- Representación de datos: Matriz de Asignación de Responsabilidades

		Roles				
		Gerente del Producto	Director del Proyecto	Scrum Master	Interesados muy influyentes	Interesados poco influyentes
Tareas	Acta de Constitución del Proyecto	A	R	I		
	Identificar los Interesados	A	R	I		
	Plan del Proyecto	A	R	C	C	I
	Planificación	A	R	I	C	I
	Dirigir el equipo		I	R		

*Tabla 14. Matriz de Asignación de Responsabilidades.*

### *Planificar la Gestión de las Comunicaciones*

Debe identificarse las necesidades de información de los interesados del proyecto. A partir de estos datos es posible planificar qué, cómo y cuándo informar las informaciones importantes del proyecto. Se debe establecer responsables para comunicar ciertas informaciones, y métodos para establecer el modo más efectivo.

El equipo de desarrollo mayormente ha de comunicarse en modalidad cara-a-cara por el enfoque ágil considerado en esta propuesta. Las demás comunicaciones formales pueden documentarse para fines de controlar el flujo de información entre los involucrados del proyecto.

Herramientas:

- Juicio de expertos
- Análisis de requisitos de comunicación
- Tecnología de la comunicación
- Habilidades interpersonales y de equipo
- Reuniones

### *Planificar la Gestión de los Riesgos*

Los riesgos son parte inherente de cada proyecto, y más en enfoques adaptativos para productos de Minería de Datos en los cuales se persigue predecir el comportamiento de clientes a partir de investigaciones que no garantizan el desarrollo un modelo óptimo por las informaciones a las cuales se tiene acceso, los recursos disponibles, las condiciones del mercado, entre otros factores.

Tanta incertidumbre amerita en una fase de Iniciación la gestión de riesgos que no está considerada en CRISP-DM.

Herramientas:

- Juicio de expertos
- Análisis de los datos
- Reuniones

### *Planificar la Gestión de las Adquisiciones*

Otro proceso no tomado en cuenta en CRISP-DM y el modelo ágil Scrum es Planificar la Gestión de las Adquisiciones. No planificar las adquisiciones/contrataciones de recursos externos que amerite el proyecto puede representar una mala gestión. El avance de los trabajos puede verse comprometido por no contar con el soporte de entidades que no pertenecen al equipo de trabajo.

Debe establecerse un plan de aprovisionamiento flexible a las necesidades del equipo de trabajo que permita obtener el soporte externo en el momento preciso.

Herramientas:

- Juicio de expertos
- Recopilación de datos
- Análisis de datos
- Análisis de selección de proveedores
- Reuniones
- Decisiones de hacer o comprar
- Establecer modelo de relación con proveedor: Partenariado o Post-Partenariado preferiblemente

### *Planificar el Involucramiento de los Interesados*

En entornos ágiles es primordial el involucramiento de los interesados en el proyecto.

En esta propuesta se incluye como una planificación inicial que ha de servir de guía para concienciar al equipo e involucrados la importancia del compromiso de cada

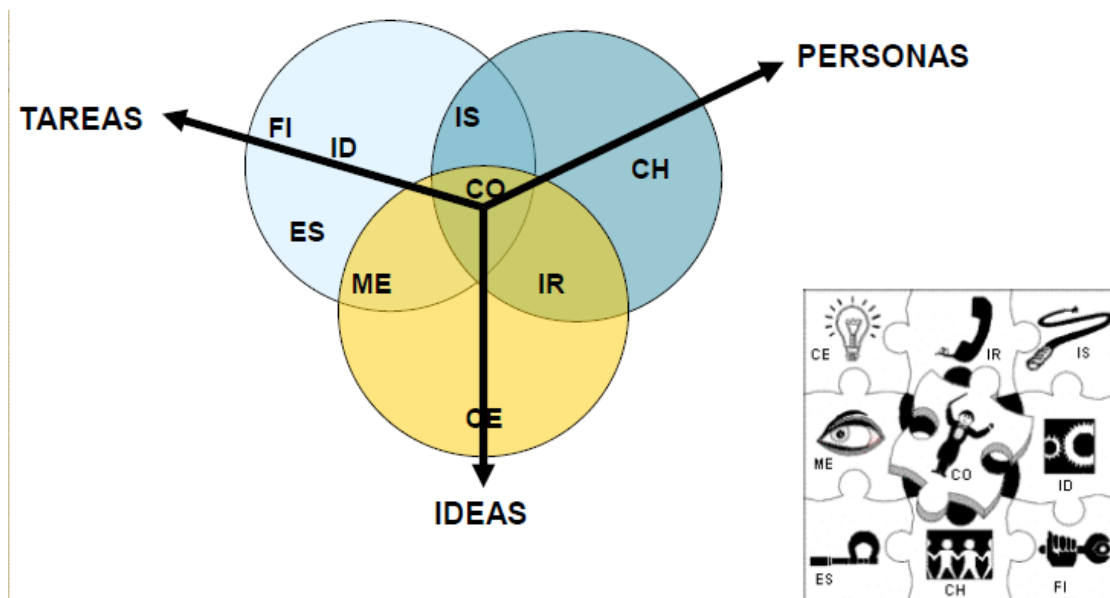
individuo. Debe plasmarse en este proceso los principios del Manifiesto Ágil para promover la colaboración.

El involucramiento en las fases iterativas tiene poco que planificar, sin embargo, se incluye este proceso para fines de resaltar la importancia que debe prestar la organización a este enfoque de trabajo colaborativo.

El Director de Proyecto es el responsable de aplicar las estrategias descritas en esta planificación.

Herramientas:

- Juicio de expertos
  - Reuniones
  - Planes de Motivación al personal
  - Orientación de los papeles en el equipo de trabajo (Belbin)
- Aplicar combinaciones de Roles que permitan conformar un equipo eficaz.

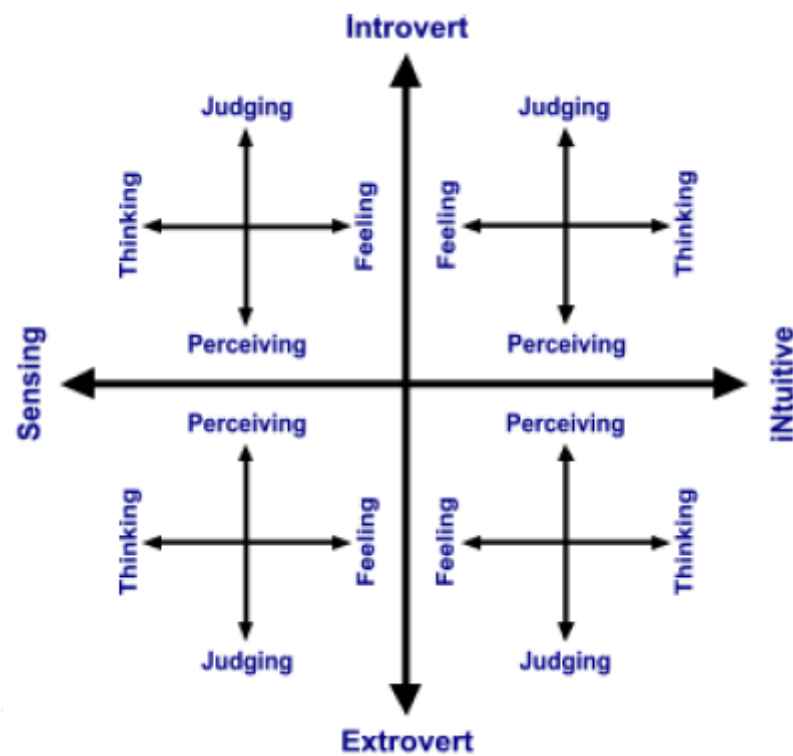


*Ilustración 9. Roles de Equipo Belbin*

Roles	
CE	Cerebro
IR	Investigador de Recursos
IS	Impulsor
CO	Coordinador
CH	Cohesionador
ID	Implementador
ME	Monitor-Evaluador
FI	Finalizador
ES	Especialista

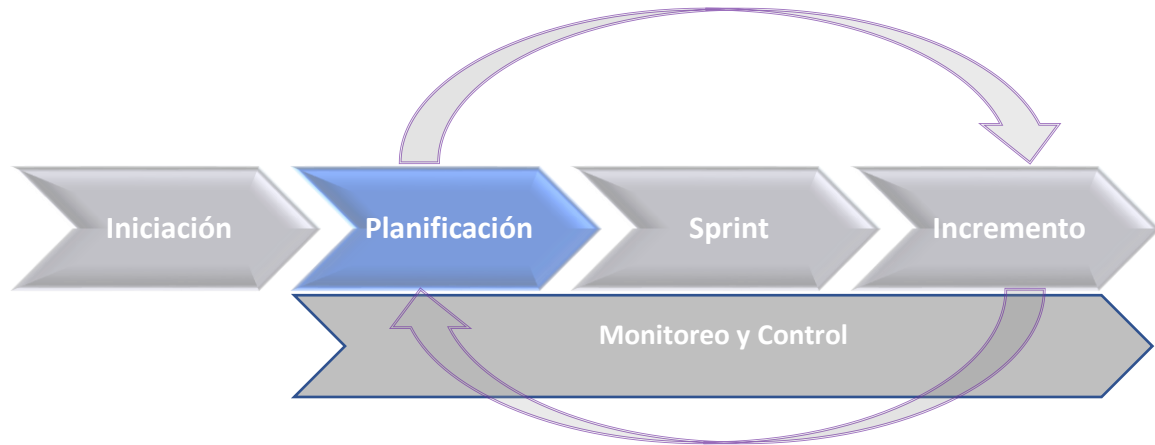
*Tabla 15. Roles de Equipo Belbin*

- Myers-Briggs Type Indicator (MBTI)  
Medición de preferencias entre los individuos del equipo.



*Ilustración 10. Indicador de tipo de Myers-Briggs MBTI*

## Fase de Planificación: Enfoque adaptativo



*Ilustración 11. Propuesta de Marco de Trabajo -Planificación-*

En el enfoque adaptativo se recomienda utilizar ciclos cortos para llevar a cabo el trabajo, revisar los resultados y adaptarse, según sea necesario. Estos ciclos proporcionan retroalimentación rápida sobre los enfoques y la idoneidad de los entregables, y generalmente se manifiestan como programación iterativa y programación a demanda.

### *EN-Determinar Metas de Minería de Datos*

Esta tarea incide en la fase de Planificación por representar las tareas de definición de objetivos de negocio y técnicos asociados.

Se considera parte esencial en la auto-organización del equipo. El Propietario del Producto es responsable de definir las metas de negocio en cada iteración, mientras que el Equipo de desarrollo establece los objetivos técnicos requeridos para cumplir con las especificaciones dadas por el Propietario del Producto.

La Pila del Sprint ha de servir como fundamento de planificación para el próximo periodo de trabajo y documentación de las metas a perseguir.

En la fase de Planificación forma parte de la sesión de trabajo Planificación del Sprint.



### *ED-Recoger Datos Iniciales*

Tarea de CRISP-DM destinada a lograr tener acceso a los datos determinados en el listado de recursos.

Debe realizarse una lista de las tablas o fuentes de datos, su ubicación y métodos para acceder a ellos. Esta información ha de servir para trabajos de futuras iteraciones que necesiten acceder a estas informaciones.

### *ED-Describir los Datos*

Tarea de CRISP-DM destinada a describir en detalle la información disponible para alcanzar los objetivos del Sprint.

Debe elaborarse un reporte que describa la información disponible, la cual debe satisfacer las metas descritas en la Pila del Sprint.

Esta tarea ha de repetirse las veces que sea necesaria.

### *ED-Explorar los Datos*

Como parte de la fase de Planificación debe examinarse los datos disponibles con el objetivo de determinar posibles patrones a simple vista antes de continuar con próximas fases.

A partir de los hallazgos encontrados es posible que sea necesario actualizar la Pila del Sprint y ajustar las metas previstas.

### *ED-Verificar la Calidad de los Datos*

Debe validarse que los datos son representativos e interpretables. Debe listarse el resultado de la evaluación para indicar si los datos cuentan con la calidad necesaria para continuar con las próximas fases.

### *Planificación del Sprint*

La Planificación del Sprint es una actividad que se toma del modelo de procesos Scrum con el objetivo de recopilar los requisitos que se deben tomar en cuenta en la próxima iteración de trabajo. Se establece el tiempo de un mes o menos con el objetivo de generar entregables de valor lo más rápido posible.

La Planificación del Sprint incluye los siguientes procesos del PMBOK y actividades de Scrum.

- **Definir las Actividades**

En este proceso se definen las actividades requeridas para crear el entregable que se acuerde con el Propietario del producto. Estas actividades serán las únicas a considerar dentro de la siguiente iteración o Sprint. El trabajo a realizar se documenta en la Pila del Sprint.

- **Estimar la Duración de las Actividades**

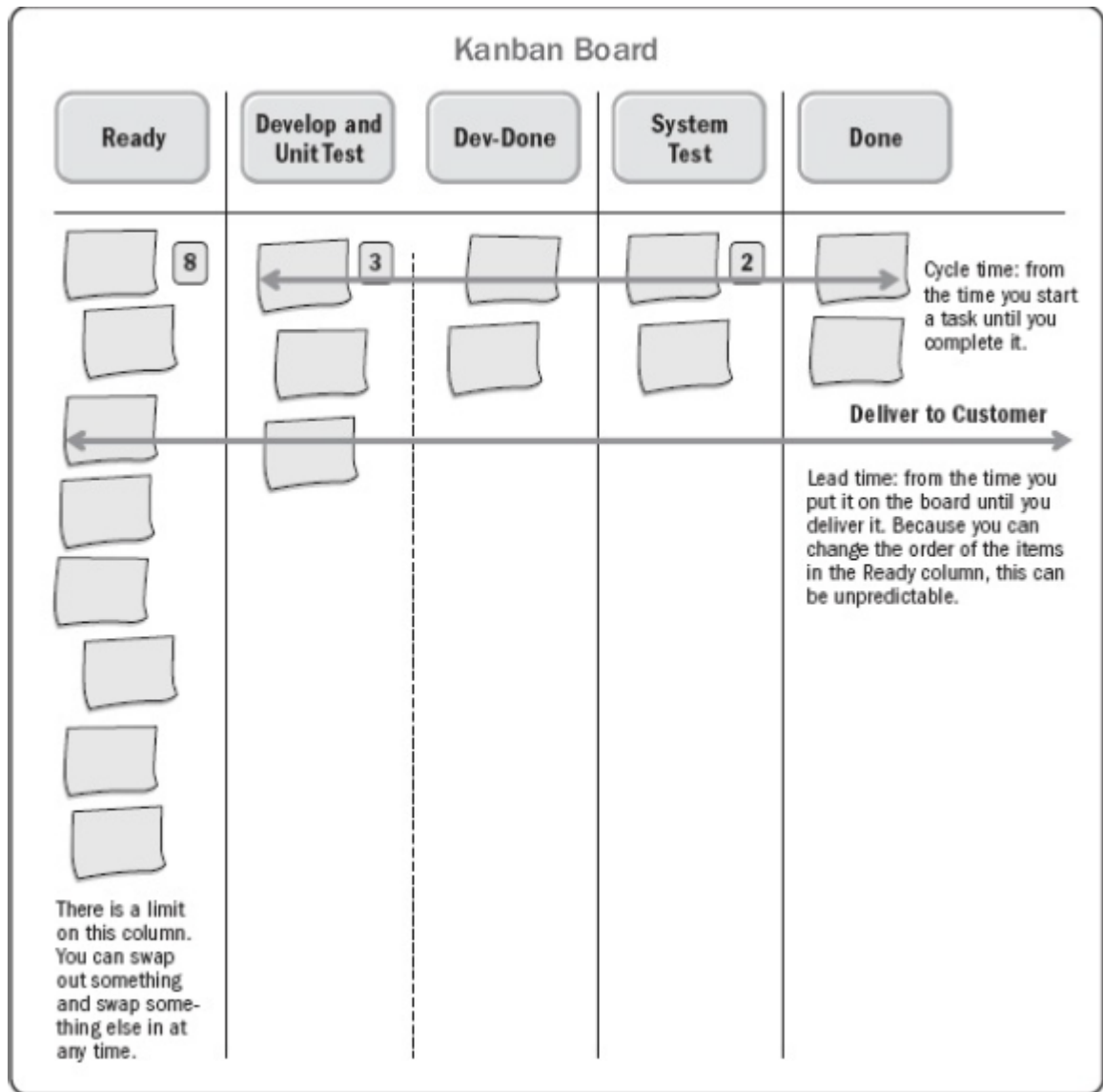
Junto a la definición del trabajo a desarrollar debe realizarse la estimación correspondiente a cada requisito. Ha de expresarse en unidades de horas.

- **Resolución de dudas**

A través de esta tarea se establece formalmente una sesión de alineación respecto a la visión que debe tener el equipo al momento de construir las especificaciones descritas en la Pila del Sprint.

Herramientas:

- Juicio de experto
- Análisis de datos
- Planificación gradual
- Reuniones
- Kanban



*Ilustración 12. Tabla Kanban, (Project Management Institute, 2017b)*

### *Planificar la Gestión del Cronograma*

Se requiere establecer en el equipo como se dará seguimiento a la planificación de trabajo y controlará el cronograma con las asignaciones y avances.

Herramientas:

- Análisis de datos
- Reuniones

### *Desarrollar el Cronograma*

Debe definirse un cronograma de trabajo con estimación global de la totalidad del proyecto con el objetivo de tener una visión general del avance del proyecto. Este plan de trabajo será actualizado en cada iteración.

La guía esencial para el equipo de trabajo en su día a día ha de establecerse en la planificación del Sprint.

Herramientas:

- Planificación ágil de liberaciones
- Método de la ruta crítica

### *Identificar los Riesgos*

Proceso parte del conjunto de las buenas prácticas recomendadas por el PMI en el PMBOK. Se incorpora en la fase de Planificación acciones que permitan al Director de Proyecto conocer los riesgos a los que se enfrenta el proyecto en cada iteración. Es de suma importancia tener identificadas los posibles detonantes de eventos que puedan hacer afectar el proyecto. También se reúne información para que el equipo del proyecto pueda responder adecuadamente a los riesgos identificados.

Herramientas:

- Tormenta de ideas
- Entrevistas
- Análisis de causa raíz
- Análisis FODA

### *Realizar el Análisis Cualitativo de Riesgos*

Debe realizarse un análisis por cada riesgo identificado que permita clasificarlos en base a su criticidad y probabilidad de ocurrencia. Realizar el Análisis Cualitativo de Riesgos establece las prioridades relativas de los riesgos individuales del proyecto para Planificar la Respuesta a los Riesgos. Se identifica un dueño del riesgo para cada riesgo, quien va a asumir la responsabilidad de planificar una respuesta adecuada al riesgo y garantizar que se implemente.

El proceso Realizar el Análisis Cualitativo de Riesgos ha de llevarse a cabo de manera regular a lo largo del ciclo de vida del proyecto, por cada iteración.

## Herramientas:

- Juicio de expertos
- Evaluación de probabilidad e impacto de los riesgos

Impacto/Consecuencia			
Probabilidad de ocurrencia	Marginal	Significativo	Crítico
Muy probable	Riesgo medio	Riesgo alto	Riesgo alto
Probable	Riesgo bajo	Riesgo medio	Riesgo alto
Improbable	Riesgo bajo	Riesgo bajo	Riesgo medio

Tabla 16. Matriz de evaluación de riesgos

## Planificar la Respuesta a los Riesgos

Debe ponderarse opciones estratégicas como respuesta a los riesgos ya identificados. El beneficio de este proceso es que identifica las formas adecuadas de abordar el riesgo general e individuales del proyecto.

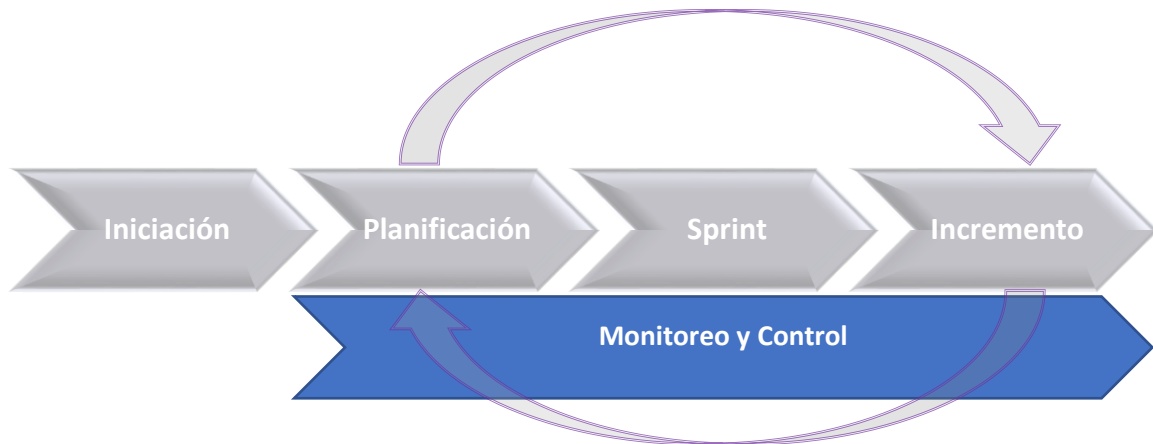
## Herramientas:

- Juicio de expertos
  - Establecer estrategias de respuesta
- Definir respuestas para cada riesgo:
- Escalar
  - Evitar
  - Transferir
  - Mitigar
  - Aceptar

Risk Identification										Qualitative Display of Most Likely Impact			Response		Monitoring and Control			Critical Issue																																			
Risk #	Status	RBS Group	Project Phase-Date Identified	Phase (pre CN, CN, or ROW)	Summary Description Threat and/or Opportunity	Detailed Description of Risk Event (Specific, Measurable, Attributable, Relevant, Timebound) [SMART]	Risk Trigger	Probability	Impact	Risk Matrix	Strategy	ACTION TO BE TAKEN Response Actions including advantages and disadvantages include date	Risk Owner	Risk Review Dates	Date, Status and Review Comments (Do not delete prior comments, therefore providing a history)	Is Risk on Critical Path?																																					
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(13)	(14)	(15)	(16)	(17)	(18)	(19)	(20)	(21)																																					
EXAMPLE	Active	Environmental	Design/PS&S	mar-05	Threat	The mitigation ratio has not been finalized and also there could be additional impacts to wetlands which would increase the amount of R/W needed for the mitigation area.	If Wetland impact is larger than 1/2 acre and ratio exceeds 4:1.	High	Very High	<table><tr><td>VH</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>H</td><td>Mo</td><td></td><td></td><td></td><td>\$</td></tr><tr><td>M</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>L</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td>VL</td><td></td><td></td><td></td><td></td><td></td></tr><tr><td></td><td>VL</td><td>L</td><td>M</td><td>H</td><td>VH</td></tr></table>	VH						H	Mo				\$	M						L						VL							VL	L	M	H	VH	Mitigation	Finalize design to identify all wetlands that are impacted. Early coordination with the outside agencies to determine mitigation ratio.	Design Leader/Enviro. mgr	2006- Dec-2	2007-Jan-2	As of Nov. 15, 2005 there are only two potential areas where there could be additional wetland impacts. As of Dec. 2, 2005 agency has initially determined that mitigation ration would be 4:1.	YES
					VH																																																
					H				Mo					\$																																							
					M																																																
L																																																					
VL																																																					
	VL	L	M	H	VH																																																
Very Low																																																					

Tabla 17. Matriz de respuesta y monitoreo a riesgos

## Monitoreo y Control: Enfoque adaptativo



*Ilustración 13. Propuesta de Marco de Trabajo -Monitoreo y Control-*

Esta propuesta de trabajo plantea una fase transversal no considerada en CRISP-DM. Se fundamenta en prácticas descritas en el PMBOK que beneficiaran al proyecto brindando la formalidad y seguimiento que desea la alta dirección en las organizaciones, sin comprometer la flexibilidad de una metodología ágil.

Las habilidades del Scrum Master son esenciales en el enfoque de esta fase, ya que el papel de este rol no representa en el proyecto un simple miembro del equipo, sino que se le confiere la responsabilidad de dirigir un equipo multidisciplinario sin descuidar los intereses de negocio que persigue la organización.

Esta fase está compuesta por una serie de procesos que inician junto a la Planificación y concluyen con los incrementos requeridos por el proyecto.

### *Dirigir y Gestionar el Trabajo del Proyecto*

Aunque es parte de la filosofía ágil contar con equipos auto-organizados, es posible que surjan situaciones no contempladas en el proyecto. No siempre se manifiesta la armonía entre los individuos que conforman los equipos de trabajo. El Scrum Master debe tener la táctica de incidir en los comportamientos de los involucrados, teniendo en cuenta sus intereses y los de la organización.

Este proceso no debe dar como resultado la creación de una figura líder autoritaria que asigne y vele por el cumplimiento fiel de las tareas, sino más bien, debe propiciar una gestión de colaboración que genere sentido de compromiso entre los miembros del equipo de trabajo. De esta manera la productividad será mayor de lo esperado.

Herramientas:

- Juicio de expertos
- Reuniones

### *Gestionar el Conocimiento del Proyecto*

La Transparencia es uno de los pilares de Scrum. El conocimiento adquirido debe ser de fácil acceso para todo el equipo y la organización. Los beneficios clave de este proceso son que el conocimiento previo de la organización se aprovecha para producir o mejorar los resultados del proyecto y que el conocimiento creado por el proyecto está disponible para apoyar las operaciones de la organización y los futuros proyectos o fases.

Herramientas:

- Juicio de expertos
- Capacitaciones
- Grupos focales
- Técnicas de gestión de creatividad

### *Gestionar la Calidad*

Gestionar la Calidad es la instancia de la planificación realizada en la fase de Iniciación. Mediante el seguimiento a políticas de calidad se incrementa la probabilidad de cumplir con objetivos de calidad tanto de la ejecución del proyecto, como de la construcción del producto. Se puede identificar procesos ineficaces que pueden impedir alcanzar las metas de negocio.

Herramientas:

- Recopilación de datos
- Análisis de datos
- Representación de datos
- Cinco ¿Por qué? (Causa raíz)
- Ciclo de mejora continua
- Causa-efecto (Ishikawa)
- 5H 2H

### *Adquirir Recursos*

Existen proyectos de Minería de Datos en los que es necesario adquirir recursos no disponibles en inicios del proyecto. El seguimiento a los planes de aprovisionamiento permite que el equipo tenga a mano lo necesario en el momento preciso. El Director de Proyecto debe velar por la asignación correcta de los recursos, seguimiento al calendario sobre su disponibilidad, previsión, etc.

Es un proceso que puede realizarse en todas las fases del desarrollo del producto según lo requiera el proyecto.

Herramientas:

- Toma de decisiones
- Preasignación

### *Desarrollar el Equipo*

No debe descuidarse el proceso de mejorar las competencias, interacción de los miembros del equipo y el ambiente general del equipo para lograr un mejor desempeño del proyecto. Scrum persigue este ideal, sin embargo, no se logra si no existe la disposición de un líder que motive al personal a dar lo mejor de sí cada día. El Scrum Master debe procurar que los miembros del equipo desarrollen sus habilidades interpersonales y competencias profesionales.

Herramientas:

- Co-ubicación
- Equipos
- Reconocimiento y recompensas
- Motivación
- Capacitación
- Evaluaciones individuales y de equipo
- Reuniones

### *Dirigir al Equipo*

En esta propuesta, una de las principales funciones del Scrum Master es Dirigir al Equipo. Debe dar seguimiento al desempeño de los miembros del equipo, proporcionar retroalimentación, resolver conflictos y gestionar cambios en el modo de trabajo para optimizar el desempeño del proyecto.

El valor del Scrum Master en este enfoque de trabajo no reside en su posición, sino en su habilidad de sacar lo mejor de los demás.

Por otro lado, trabajar en un entorno ágil de desarrollo de productos complejos hace que el rol del Director de Proyecto cambie el paradigma como agente centralizado



que dictamina el curso del proyecto, y lo convierte en un colaborador que sirva al equipo y a la gestión.

La combinación entre un modelo predictivo y ágil requiere la estructuración del siguiente equipo de trabajo:

- Director de proyecto: En proyectos de alta variabilidad existen complicaciones que una sola persona no puede gestionar. Se designa la responsabilidad al Director de Proyecto de velar por la planificación general, seguimiento y control de los recursos, costos y riesgos que pudieran afectar al proyecto. El Director del proyecto debe encajar lo más posible en la filosofía ágil, delegando lo más que se pueda. Ha de ser el enlace entre el Equipo de desarrollo, la alta gerencia, el Gerente del producto y entidades externas.
- Scrum Master: Debe facilitar el trabajo en equipo coordinando las actividades entre el Propietario del producto y Equipo de desarrollo. Junto al Propietario del producto es quien coordina junto al Director del proyecto las medidas de control que sean necesarias aplicar durante el desarrollo del proyecto.
- Gerente del producto: Ejecutivo con visión de negocio que no es parte de departamentos operativos de la empresa como lo es el Propietario del producto.
- Propietario del producto: Representante del negocio que domina las operaciones de la empresa. Debe administrar los requisitos del proyecto.
- Equipo de desarrollo: Personal que debe elaborar el producto de valor para la organización.

Herramientas:

- Habilidades interpersonales y de equipo

### *Gestionar las Comunicaciones*

El Director de Proyecto debe velar por el efectivo flujo de información entre el equipo y hacia la alta gerencia. Debe garantizar que la recopilación, distribución, almacenamiento necesario, monitoreo y disposición final de la información del proyecto sea oportunos y adecuados.

Las necesidades del proyecto dictarán en qué medida debe gestionarse el flujo de estas informaciones.

Herramientas:

- Tecnología de la comunicación
- Métodos de comunicación
- Habilidades de comunicación
- Generación de informes del proyecto
- Habilidades interpersonales y de equipo
- Reuniones

### *Implementar la Respuesta a los Riesgos*

A través de este proceso se aplica las respuestas que se han planificado respecto a los riesgos identificados.

Con el objetivo de disminuir las amenazas y aprovechar las oportunidades, el Director de Proyecto asume la responsabilidad de implementar las respuestas idóneas frente a los riesgos que pueden afectar al proyecto.

Herramientas

- Juicio de expertos
- Habilidades interpersonales y de equipo

### *Efectuar las Adquisiciones*

Este proceso representa la ejecución del plan de aprovisionamiento preparado para contar con los recursos del proyecto en el momento preciso. Efectuar adquisiciones efectivas según un plan sistemático de aprovisionamiento da como resultado concretar negociaciones con proveedores calificados y homologados bajo criterios de calidad previamente establecidos.

La necesidad del proyecto determina si aplica la implementación de este proceso y el nivel de adaptación en el marco ágil de trabajo.

Herramientas:

- Juicio de expertos
- Conferencias de oferentes
- Análisis de datos
- Habilidades interpersonales y de equipos

### *Gestionar la Participación de los Interesados*

A pesar de contar con un equipo de trabajo auto-organizado, debe gestionarse la integración de los interesados en el desarrollo del proyecto. Esta gestión se incluye en esta propuesta con el objetivo de dar las herramientas necesarias al líder del

proyecto para fomentar el apoyo y minimizar las posibles resistencias de los individuos por cambios que surjan en el desarrollo del producto.

Herramientas:

- Juicio de expertos
- Habilidades de comunicación
- Habilidades interpersonales y de equipo
- Reuniones

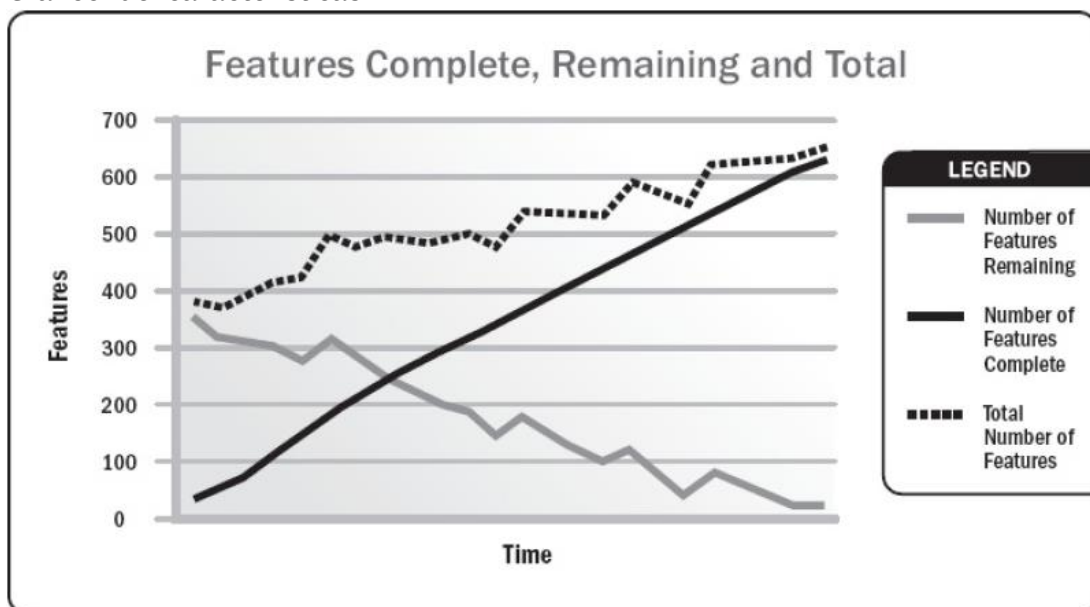
### *Monitorear y Controlar el Trabajo del Proyecto*

Para evitar que el equipo de trabajo pierda el enfoque, el Director de Proyecto debe realizar tareas de monitoreo y control sobre los trabajos planificados para los Sprints. De forma transparente, todo el equipo debe conocer el estado actual del proyecto y medidas adoptadas para eficientizar el desempeño.

Este proceso debe permitir al equipo tener una visión a futuro del proyecto respecto al trabajo general a realizar.

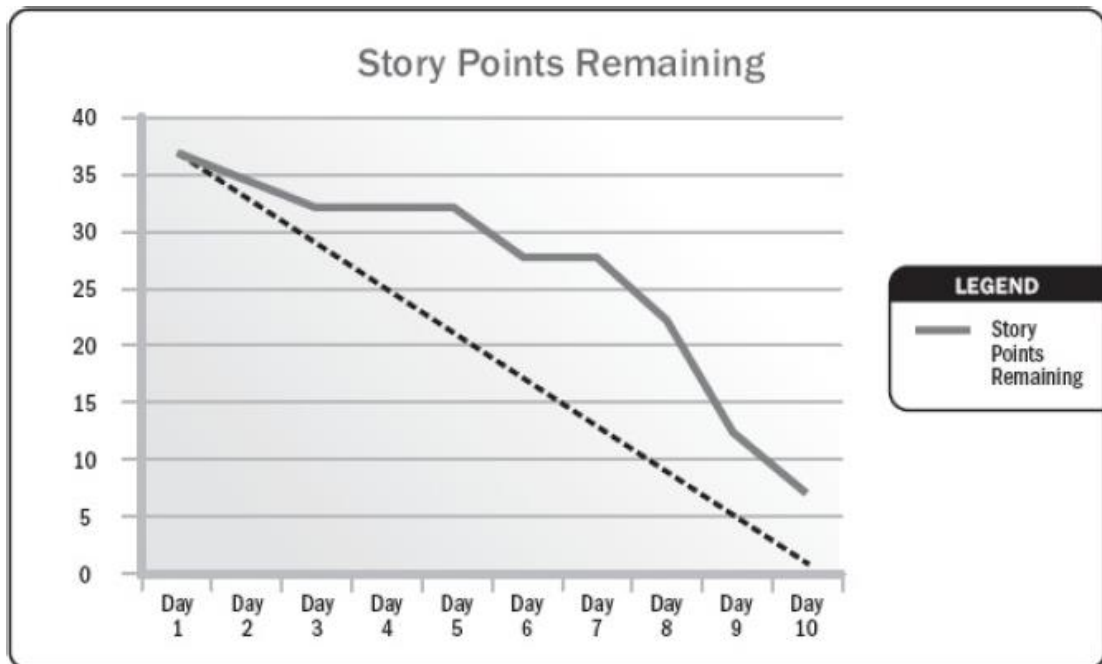
Herramientas:

- Juicio de expertos
- Análisis de datos
- Reuniones
- Gráfico de características



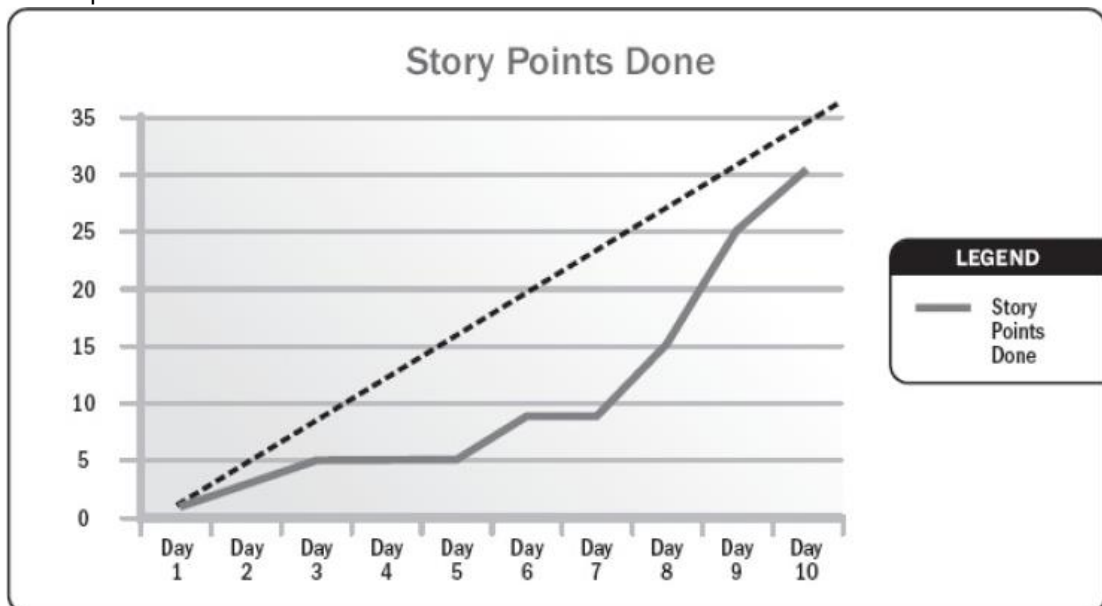
*Ilustración 14. Gráfico de características, (Project Management Institute, 2017b)*

- BurnDown Charts



*Ilustración 15. Diagrama BurnDown, (Project Management Institute, 2017b)*

- BurnUp Charts



*Ilustración 16. Diagrama BurnUp, (Project Management Institute, 2017b)*

El BurnUp Chart muestra el trabajo completado comparado al total de trabajo que se espera realizar por iteraciones.

### *Realizar el Control Integrado de Cambios*

Para cumplir con los tres pilares del marco de procesos Scrum: transparencia, inspección y adaptación, es vital realizar un control integrado de cambios que permita disponer de la trazabilidad de la documentación y artefactos del producto en el momento que se requiera.

El proceso de cambios debe ser flexible si lo comparamos con metodologías tradicionales de desarrollo de software, que presentan políticas burocráticas para administrar los cambios en el proyecto. En nuestro caso, el equipo auto-organizado acepta los cambios fácilmente que solicite el Propietario del Producto si está dentro de los objetivos del proyecto.

Siempre debe existir el historial de cambios como insumo para que el Director de Proyecto pueda evaluar posibles desviaciones frente a los objetivos iniciales, de manera tal que pueda aplicar las medidas ideales para reenfocar al equipo de ser necesario.

Herramientas:

- Juicio de expertos
- Herramientas automatizadas para control de cambios
- Reuniones

### *Controlar el Cronograma*

Debe determinarse el estado actual del cronograma del proyecto mediante la comparación de la cantidad total de trabajo entregado y aceptado con respecto a las estimaciones de trabajo estimado para el Sprint y los requisitos pendientes de la Pila de Producto.

La alta gerencia usualmente quiere saber cuál es la fecha de término de los proyectos. A medida en que el Equipo de desarrollo avanza puede determinarse un promedio de historias de usuarios por iteración, el Director de Proyecto puede predecir que tanto tiempo tomaría concluir el proyecto.

Puede utilizarse la planificación original de trabajo e ir alimentándola con los avances realizados por el equipo, de manera tal que pueda reflejarse la evolución y curso que sigue el proyecto.

Herramientas:

- Optimización de recursos
- Ajustes de adelantos y retrasos

- Comprensión del cronograma

### *Controlar los Costos*

El Director de Proyecto debe monitorear constantemente los gastos en los cuales incurre el proyecto. Debe implementar las estrategias que ameriten para que se mantenga el presupuesto y le permita lograr los objetivos estratégicos del proyecto.

Es aceptable que el costo descrito en la línea base definida en la fase de Iniciación difiera respecto a la ejecución real, sin embargo, el Director de Proyecto debe asegurar que la desviación sea mínima para no perjudicar las reservas que provee la Organización.

Herramientas:

- Juicio de expertos
- Análisis de tendencia
- Análisis de reserva
- Análisis del Valor Ganado
- Análisis de variación: Valor Actualizado Neto (V.A.N)

$$VAN = \frac{FC_1}{1+i} + \frac{FC_2}{(1+i)^2} + \dots + \frac{FC_n}{(1+i)^n} - FC_0$$

- Análisis de variación: Tasa Interna de Retorno (T.I.R)

$$FC_0 = \frac{FC_1}{1+i} + \frac{FC_2}{(1+i)^2} + \dots + \frac{FC_n}{(1+i)^n}$$

### *Controlar la Calidad*

El enfoque de este proceso debe ser para esta propuesta, monitorear y controlar que las actividades de gestión de calidad se realizan como se ha previsto. Debe evaluarse el desempeño del trabajo y asegurar que los involucrados cumplan con sus responsabilidades.

Está orientada a controlar la calidad de la ejecución de los procesos del proyecto y el uso correcto de los recursos. Mide también la integridad, cumplimiento y adecuación para el uso de los entregables que genera cada iteración de trabajo.

Herramientas:

- Recopilación de datos: Listas de verificación

- Análisis de datos: Revisión de desempeño
- Inspección
- Indicadores de Calidad

### *Controlar los Recursos*

Es responsabilidad del Director de Proyecto asegurar que los recursos asignados al proyecto están disponibles como ha sido planificado, y liberarlos cuando ya no son necesarios. Este proceso es importante ya que puede significar un retraso de trabajo si el proyecto no dispone de los recursos apropiados en el momento que lo amerite.

Herramientas:

- Análisis de alternativas
- Análisis costo-beneficio

### *Monitorear los Riesgos*

Debe supervisarse la implementación de los planes acordados de respuesta a los riesgos, hacer seguimiento a los riesgos identificados y evaluar la efectividad del proceso de gestión de los riesgos.

Este proceso sirve para determinar si las respuestas a los riesgos son efectivas, el nivel de riesgo general del proyecto ha cambiado, si el estado de los riesgos identificados ha cambiado, si el enfoque de gestión de riesgos sigue siendo adecuado, si son seguidas las políticas y procedimientos de gestión de riesgos, etc.

Herramientas:

- Análisis de los datos
- Reuniones

### *Controlar las Adquisiciones*

Debe gestionarse las adquisiciones, contrataciones con proveedores, aplicar cambios y correcciones según corresponda y cerrar contratos.

El Director de Proyecto debe asegurar que los acuerdos con empresas externas y el producto de las negociaciones son efectuados bajo el cumplimiento de las obligaciones contractuales.

Herramientas:

- Juicio de expertos
- Inspección

### *Monitorear el Involucramiento de los Interesados*

Siendo parte del equipo, el Director de Proyecto debe monitorear las relaciones de los interesados del proyecto y adaptar las estrategias para los fines de incentivar dichas relaciones.

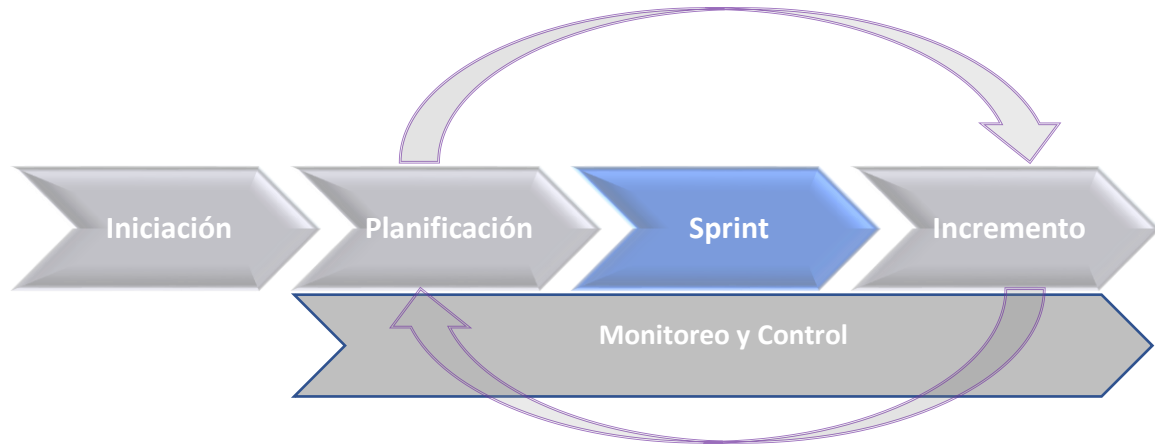
La dinámica de trabajo que presenta Scrum propicia el desarrollo de relaciones entre los miembros del equipo y demás involucrados. Reuniones diarias, trabajo en equipo, retrospectivas, transparencia y filosofía de trabajo permiten que el involucramiento de los interesados sea mayor que otros modelos de gestión de proyectos.

Herramientas:

- Análisis de los datos: Análisis causa raíz
- Análisis de los datos: Análisis de interesados
- Toma de decisiones
- Habilidades de comunicación
- Habilidades interpersonales y de equipo
- Reuniones



## Fase Sprint: Enfoque adaptativo



*Ilustración 17. Propuesta de Marco de Trabajo -Sprint-*

La propuesta descrita en este estudio recomienda implementar la fase de Sprint que presenta el modelo de procesos Scrum. Esta fase incorpora las actividades de ejecución asociadas a Minería de Datos que tiene CRISP-DM como medida de adaptación.

### *PD-Seleccionar los Datos*

El Equipo de desarrollo debe listar los datos a ser utilizados y motivos de la toma de decisión. Considerar técnicas de muestreo. Documentar las razones de la inclusión o exclusión de los atributos.

### *PD-Limpiar los Datos*

Corregir, remover o ignorar los errores en los datos. Documentar como tratar los valores especiales y sus respectivos significados.

### *PD-Construir los Datos*

Construir atributos derivados de otros, completar nuevos registros, o transformar valores a partir de atributos existentes.

### *PD-Integrar los Datos*

Comprobar si las herramientas de integración pueden integrar las diversas fuentes de entrada como se requiere. Integrar fuentes de datos.

### *PD-Formatear los Datos*

Dar sentido a los datos a utilizar. Cambios sintácticos para satisfacer las exigencias de la herramienta específica de modelado.

### *CM-Seleccionar Técnica de Modelado*

Documentar las técnicas a emplear.

### *CM-Generar Diseños de Pruebas*

Definir los pasos necesarios para ejecutar las pruebas (número de iteraciones, métricas objetivo, etc.). Preparar datos requeridos para la prueba.

### *CM-Construir Modelo*

Crear los modelos requeridos, describirlos, interpretarlos y documentar las dificultades para comprender su significado.

Describir el comportamiento del modelo y su interpretación. Expresar conclusiones sobre los modelos construidos.

### *Reunión diaria (Daily Scrum)*

En adición a las preguntas descritas en el modelo Scrum, incorporar a las siguientes:

- ¿Qué necesitamos para avanzar el trabajo priorizado por el Propietario del Producto?
- ¿Alguien está trabajando en algo que no esté en la Pila del Sprint?
- ¿Qué recursos adicionales necesitamos para terminar como equipo?
- ¿Existen cuellos de botellas que impidan un flujo de trabajo normal?

Estas reuniones deben ser de no más de 15 minutos de pie. Solo debe participar el Equipo Scrum. No es requerida la participación del Gerente del Producto ni del Director de Proyecto. El objetivo de esta reunión es identificar las principales trabas que impiden al equipo avanzar con sus trabajos; no debe convertirse en una reunión de revisión de estado de avance.

### *CM-Evaluar Modelos*

Hacer evaluaciones de los resultados y seleccionar los mejores modelos. Interpretar los resultados en función del negocio, tanto como sea posible. Revisar los resultados junto al Propietario del Producto y conocer sus impresiones. En este momento debe validarse la fiabilidad de los resultados.

### *EP-Evaluar los Resultados*

Determinar si hay alguna condición de negocio que el modelo no cubra. Se busca con esta actividad: comprender los resultados de la minería de datos e interpretar en términos de su aplicación. El equipo debe también comprobar que los resultados son nuevos y útiles.

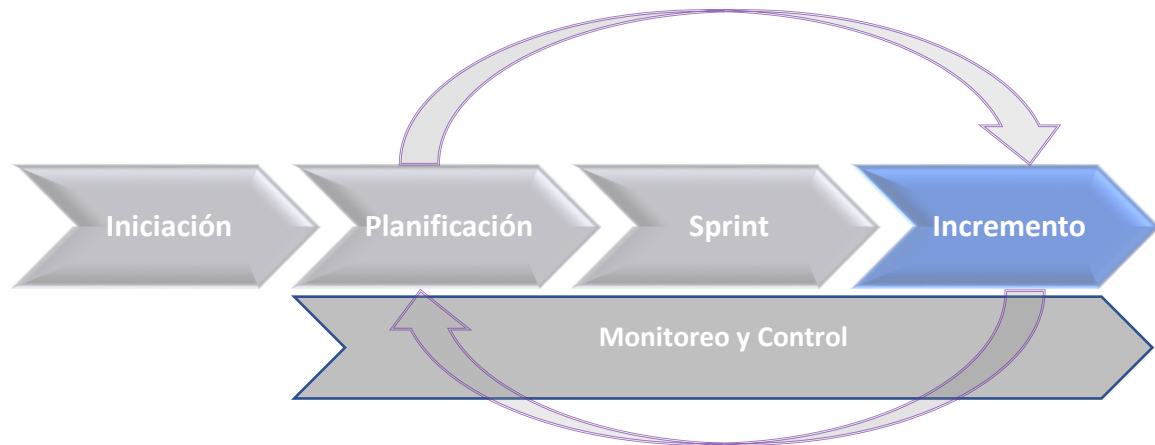
*EP-Revisar el Proceso*

El Equipo Scrum debe realizar una revisión más minuciosa de las actividades realizadas durante el Sprint con el objetivo de confirmar que no han pasado algo por alto.

Debe documentarse las acciones tomadas y evaluar cuales pueden descartarse para un próximo Sprint.

Identificar fracasos, desviaciones y acciones alternativas que puede adoptar el Equipo.

## Fase Incremento: Enfoque adaptativo



*Ilustración 18. Propuesta de Marco de Trabajo -Incremento-*

La propuesta descrita en este estudio recomienda implementar la fase de Incremento que presenta el modelo de procesos Scrum. Esta fase incorpora las actividades de ejecución asociadas a la revisión del proyecto y producto al final de cada Sprint. En esta parte del proyecto se concretan los entregables que han de dar valor al proyecto y permiten que la Organización reconozca avances tangibles del proyecto.

Al crear pequeños incrementos, probarlos y luego revisarlos, permite al equipo explorar la incertidumbre a un bajo costo en corto periodo de tiempo, reducir el riesgo y maximizar entregas con valor comercial.

### *Retrospectiva*

Una vez concluida la fase de incremento, antes de iniciar con la planificación del próximo Sprint, el equipo debe realizar reunión de retrospectiva en la cual analice cómo ha sido su manera de trabajar durante la iteración previa.

Debe evaluar los siguientes puntos:

- Cómo han funcionado sus estrategias y si deben mejorar
- Qué estrategias de colaboración desean incorporar en el próximo Sprint
- Qué se ha aprendido hasta el momento
- Qué problemas han dificultado el progreso del trabajo.

El Scrum Master es responsable de solucionar los inconvenientes expuestos por el equipo, de manera que en una próxima iteración el desempeño sea más eficiente.

El Director de Proyecto debe participar en estas sesiones para la toma de decisiones respecto al curso que deba llevar el proyecto. Es el momento ideal para actualizar la documentación de seguimiento y control para fines de preparar los reportes de estado del proyecto que servirán para el Equipo, Involucrados y alta gerencia.

Realizar esta reunión en un periodo corto de aproximadamente tres horas, o lo que demande la magnitud del proyecto.

Dentro de las sesiones de Retrospectiva se consideran las siguientes actividades propias de CRISP-DM, las cuales ameritan ser tomadas en cuenta en proyectos de este tipo:

- **EP-Determinar los Próximos Pasos**

Como parte de la sección de Incremento, el Equipo Scrum debe determinar cuáles han de ser las próximas acciones referente a los resultados obtenidos. Dependiendo de los resultados y estado general del proyecto, se puede tomar la decisión de continuar con nuevas iteraciones o dar por concluido el desarrollo del proyecto.

- **PP-Planificar Despliegue**

Si es necesario, el Equipo Scrum debe planificar como debe realizarse el despliegue de los cambios requeridos por los resultados derivados del Sprint. Las tareas de despliegue pueden delegarse a otro equipo de trabajo dentro de la organización o asumirse por el Equipo de desarrollo. Todo depende de la estructura de los departamentos de TI de la empresa.

- **PP-Revisar Proyecto**

La revisión del proyecto debe resaltar la experiencia obtenida en cada fase de la iteración. Esta actividad permite identificar los errores cometidos y las buenas prácticas que debe continuar realizando el equipo.

#### *PP-Planificar Monitoreo y Mantenimiento*

El Equipo de desarrollo debe documentar las especificaciones de Monitoreo y Mantenimiento necesarios para poder obtener los beneficios evaluados durante el desarrollo del proyecto.

Los despliegues en ambiente productivo deben ser monitoreados para poder confirmar el éxito del proyecto.

#### *PP-Elaborar el Reporte Final*

El Scrum Master debe documentar las lecciones aprendidas por el equipo, de manera tal que sirva a la organización como guía para futuros proyectos o Sprints del mismo.

El informe debe contener un resumen del producto construido y los resultados obtenidos de los modelos.

#### *Validar el Alcance*

En esta fase se incorpora a partir del PMBOK, el proceso de Validar el Alcance con la finalidad de documentar formalmente la aceptación de los entregables del proyecto que se han completado.

Este proceso forma parte de la fase de Incremento para no incluir complejidad al ciclo continuo propio de entornos adaptativos.

Tanto el Director de Proyecto, el Scrum Master, el Gerente del Producto y Propietario del Producto deben realizar esta validación.

Herramientas:

- Inspección
- Toma de decisiones

#### *Controlar el Alcance*

En esta fase se incorpora a partir del PMBOK, el proceso de Controlar el Alcance con la finalidad de monitorear el estado del alcance del proyecto.

La responsabilidad de llevar a cabo este monitoreo es del Director de Proyecto. Se puede auxiliar del Gerente del Producto y Propietario del Producto si hay un cambio de objetivos que deba considerarse para próximas iteraciones. Esto permitirá que sean aplicados los ajustes necesarios en próximas iteraciones o despliegues.

Herramientas:

- Análisis de datos: Análisis de variación
- Análisis de datos: Análisis de tendencias

# Capítulo 5

## Conclusiones y líneas futuras

“Si no podemos influir positivamente en la gestión, nuestro trabajo es solo trabajo,  
no un trabajo productivo”

## 5. Conclusiones y líneas futuras

### 5.1 Conclusiones

Con el análisis que realizan otros investigadores y el presente trabajo, podemos ver como las tecnologías disruptivas están cambiando rápidamente el enfoque de gestión de proyectos al traer consigo nuevos paradigmas de desarrollo de software, dirección de equipos de trabajo y creación de soluciones empresariales que garanticen el crecimiento de las organizaciones en el mercado. La alta competitividad por la aplicación de estrategias en base al conocimiento a partir de los datos impulsa a las empresas a adoptar una cultura de trabajo en la dirección de sus proyectos que promuevan la colaboración y optimice el desempeño del personal.

Estamos acostumbrados a reconocer las métricas de tiempo, costo, alcance y calidad en la dirección de proyectos como los factores más importantes para definir el éxito de un proyecto. En adición a estas métricas, podemos ver también que el éxito del proyecto debe medirse teniendo en cuenta el logro de los objetivos del proyecto.

Gran parte del éxito de un proyecto de software se basa en su forma de trabajo.

Mediante el análisis del modelo de procesos CRISP-DM para la dirección de proyectos de minería de datos que benefician a las organizaciones a captar clientes, podemos destacar que considera las principales actividades que conlleva completar proyectos de esta clase. Sin embargo, presenta debilidades en su implementación por no ofrecer tareas que apoyen a los directores en sus tareas de gestión. En adición a esto, no promueve la creación de un entorno colaborativo para sacar el mayor provecho al equipo de trabajo.

En otro orden, a partir del análisis sobre el uso de metodologías tradicionales predictivas, podemos descartar, al menos de manera exclusiva, su implementación como guía de trabajo para la dirección de proyectos de *Machine Learning*. Esto, debido a la alta interacción entre equipos y flexibilidad que requieren estos proyectos. Una metodología con ciclo de vida en Cascada no representa una alternativa factible sobre la cual crear soluciones a partir de requisitos volátiles de negocio.

De igual modo, se estudia la aplicabilidad de las metodologías ágiles en estos proyectos. Su enfoque y principios encajan en gran medida con las necesidades de gestión y aplicación de técnicas de investigación en un entorno en el cual el personal de negocios no cuenta con todos los requisitos al inicio del proyecto. Estas metodologías contribuyen a crear un entorno en el que prolifera la cultura de adopción al cambio y permite la colaboración entre equipos multidisciplinares. En



este estudio se analiza específicamente el modelo de procesos Scrum, el cual ha ganado popularidad últimamente para la dirección de proyectos como alternativa al uso de CRISP-DM por la flexibilidad que representa.

Si bien es cierto que Scrum presenta un modelo fácilmente entendible y práctico para desarrollar proyectos de análisis de datos, no ha sido concebido para esto; por tanto, carece de fases y actividades destinadas a extraer conocimientos a partir de grandes volúmenes de información.

La realización de este trabajo ha permitido identificar las prácticas que recomienda el Project Management Institute en su guía de conocimientos PMBOK y la guía de prácticas ágiles, también del PMI.

Por último, se ha desarrollado un marco de trabajo que permita combinar las prácticas que recomienda el PMI, las herramientas que reconocen otros autores en la literatura para la dirección de proyectos, la visión de las metodologías prescriptivas, el ciclo de vida de desarrollo de Scrum y su enfoque de trabajo colaborativo.

Este marco de trabajo consiste en una etapa con características de metodologías predictivas para las tareas de poca variabilidad, una serie de tres fases con un enfoque adaptativo para el desarrollo del producto, y una fase transversal que permita realizar las tareas de control y seguimiento al proyecto.

Estas recomendaciones están diseñadas para permitir a los miembros del equipo desempeñar sus funciones en un entorno colaborativo, transparente y con las herramientas necesarias para alcanzar las metas estratégicas de negocio.

## 5.2 Líneas futuras

La investigación aplicada en este trabajo es de tipo No Experimental de carácter Exploratorio.

A partir de este estudio, el cual ha permitido definir un marco de trabajo basado en prácticas reconocidas en la dirección de proyectos, se sugiere continuar con las siguientes líneas de investigación:

- Realizar mediciones de campo que permitan evaluar con más precisión la efectividad de los procesos que se han adoptado en el marco de trabajo propuesto. De esta manera ha de confirmarse su eficacia para las tareas de dirección de proyectos de minería de datos con alta variabilidad en sus requisitos.
- Estudiar otros modelos de trabajo conocidos como los son: SEMMA, Catalyst, KDD, Agile KDD, entre otros. Analizar proyectos que hayan seguido estas guías de trabajo para su gestión e identificar cuáles han sido los factores que han contribuido a su éxito o fracaso. El resultado de este análisis puede utilizarse para comparar con CRISP-DM y las prácticas recomendadas en el marco de trabajo propuesto en este estudio.

# Fuentes consultadas

- Ahmed, S., & Danti, A. (2016). Effective Sentimental Analysis and Opinion Mining of Web Reviews Using Rule Based Classifiers (pp. 171–179). Springer, New Delhi.  
[https://doi.org/10.1007/978-81-322-2734-2\\_18](https://doi.org/10.1007/978-81-322-2734-2_18)
- Bandyopadhyay, S., & Pal, S. K. (2007). Genetic Algorithms and Web Intelligence. In *Classification and Learning Using Genetic Algorithms* (pp. 243–276). Berlin, Heidelberg: Springer Berlin Heidelberg. [https://doi.org/10.1007/3-540-49607-6\\_10](https://doi.org/10.1007/3-540-49607-6_10)
- Boyd, R. L., & Pennebaker, J. W. (2017). Language-based personality: a new approach to personality in a digital world. *Current Opinion in Behavioral Sciences*, 18, 63–68.  
<https://doi.org/10.1016/J.COBEHA.2017.07.017>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (n.d.). Step-by-step data mining guide. Obtenido de <https://www.the-modeling-agency.com/crisp-dm.pdf>
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). Step-by-step data mining guide. *The CRISP-DM Consortium*. Obtenido de <https://www.the-modeling-agency.com/crisp-dm.pdf>
- Chen, A., Lu, Y., & Wang, B. (2017). Customers' purchase decision-making process in social commerce: A social learning perspective. *International Journal of Information Management*, 37(6), 627–638. <https://doi.org/10.1016/J.IJINFOMGT.2017.05.001>
- Cios, K. J., & Kurgan, L. A. (2005). Trends in Data Mining and Knowledge Discovery. In *Advanced Techniques in Knowledge Discovery and Data Mining* (pp. 1–26). London: Springer London. [https://doi.org/10.1007/1-84628-183-0\\_1](https://doi.org/10.1007/1-84628-183-0_1)
- Crisp, C., & McKenna, D. (2016). *The Art of Scrum: How Scrum Masters Bind Dev Teams and Unleash Agility* (Vol. 4). [https://doi.org/10.1007/978-1-4842-2277-5\\_10](https://doi.org/10.1007/978-1-4842-2277-5_10)
- Gaurav, K., & Kumar, P. (2017). Consumer Satisfaction Rating System Using Sentiment Analysis (pp. 400–411). Springer, Cham. [https://doi.org/10.1007/978-3-319-68557-1\\_35](https://doi.org/10.1007/978-3-319-68557-1_35)
- Hu, Y., Du, J., Zhang, X., Hao, X., Ngai, E. W. T., Fan, M., & Liu, M. (2013). An integrative framework for intelligent software project risk planning. *Decision Support Systems*, 55(4), 927–937. <https://doi.org/10.1016/J.DSS.2012.12.029>
- Huang, Y., Li, C., Wu, J., & Lin, Z. (2018). Online customer reviews and consumer evaluation: The role of review font. *Information & Management*, 55(4), 430–440.  
<https://doi.org/10.1016/J.IM.2017.10.003>
- Ishikiriya, C. S., Miro, D., & Gomes, C. F. S. (2015). Text Mining Business Intelligence: A small sample of what words can say. *Procedia Computer Science*, 55, 261–267.  
<https://doi.org/10.1016/J.PROCS.2015.07.044>
- Kashyap, P. (2017). Innovation, KPIs, Best Practices, and More for Machine Learning. In *Machine Learning for Decision Makers* (pp. 261–279). Berkeley, CA: Apress.  
[https://doi.org/10.1007/978-1-4842-2988-0\\_7](https://doi.org/10.1007/978-1-4842-2988-0_7)
- Kosinski, M., Stillwell, D., & Graepel, T. (2013). Private traits and attributes are predictable

- from digital records of human behavior. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5802–5805.  
<https://doi.org/10.1073/pnas.1218772110>
- Kurada, R. R., & Kanadam, K. P. (2018). Sentimental Analysis on Cognitive Data Using R (pp. 15–35). Springer, Singapore. [https://doi.org/10.1007/978-981-10-6653-5\\_2](https://doi.org/10.1007/978-981-10-6653-5_2)
- Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700–710. <https://doi.org/10.1016/J.IJINFOMGT.2016.04.013>
- Mahmoodi, J., Leckelt, M., van Zalk, M., Geukes, K., & Back, M. (2017). Big Data approaches in social and behavioral science: four key trade-offs and a call for integration. *Current Opinion in Behavioral Sciences*, 18, 57–62.  
<https://doi.org/10.1016/J.COBEHA.2017.07.001>
- Marbán, O., Menasalvas, E., & Fernández-Baizán, C. (2008). A cost model to estimate the effort of data mining projects (DMCoMo). *Information Systems*, 33(1), 133–150.  
<https://doi.org/10.1016/J.IS.2007.07.004>
- Marbán, O., Segovia, J., Menasalvas, E., & Fernández-Baizán, C. (2009). Toward data mining engineering: A software engineering approach. *Information Systems*, 34(1), 87–107.  
<https://doi.org/10.1016/J.IS.2008.04.003>
- Muntean, M., & Surcel, T. (2013). Agile BI – The Future of BI. *Informatica Economică*, 17(3).  
<https://doi.org/10.12948/issn14531305/17.3.2013.10>
- Netzer, O., Feldman, R., Goldenberg, J., & Fresko, M. (2012). Mine Your Own Business: Market-Structure Surveillance Through Text Mining, 31(3), 521–543.  
<https://doi.org/10.1287/mksc.1120.0713>
- Pinheiro, C. A. R., & Helfert, M. (2010). Neural Network and Social Network to enhance the customer loyalty process. In *Innovations and Advances in Computer Sciences and Engineering* (pp. 91–96). Dordrecht: Springer Netherlands. [https://doi.org/10.1007/978-90-481-3658-2\\_16](https://doi.org/10.1007/978-90-481-3658-2_16)
- Project Management Institute. (2017a). *A guide to the project management body of knowledge (PMBOK® guide)*. *Project Management Journal* (Vol. 40).  
<https://doi.org/10.1002/pmj.20125>
- Project Management Institute. (2017b). *Agile Practice Guide*. *Project Management Journal*.
- Pytel, P., Britos, P., & Garcia-Martinez, R. (2012). A Proposal of Effort Estimation Method for Information Mining Projects Oriented to SMEs. *LNBIP*, 139, 58–74. Obtenido de <http://sistemas.unla.edu.ar/sistemas/gisi/papers/LNBIP-139-pag-58-74.pdf>
- Sayad, S. (s.f.). Data Mining Map. Retrieved July 9, 2018, from <http://www.saedsayad.com/>
- Schwaber, K., & Sutherland, J. (2016). La guía de Scrum. *Scrum.Org*, 5. Obtenido de <http://creativecommons.org/licenses/by-sa/4.0/legalcode>  
<http://creativecommons.org/licenses/by-sa/4.0/legalcode>
- Sharma, S., & Osei-Bryson, K.-M. (2009). Framework for formal implementation of the business understanding phase of data mining projects. *Expert Systems with Applications*,

362), 4114–4124. <https://doi.org/10.1016/J.ESWA.2008.03.021>

Sliger, M., & Consulting, S. (2008). Agile Project Management and the PMBOK ® Guide. *PM Network*, 1–7. Obtenido de [http://www.educause.edu/visuals/shared/pd/Agile\\_project\\_management.pdf](http://www.educause.edu/visuals/shared/pd/Agile_project_management.pdf)

Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54. <https://doi.org/10.1177/0261927X09351676>

Tiwari, A., Hadden, J., & Turner, C. (2010). A New Neural Network Based Customer Profiling Methodology for Churn Prediction (pp. 358–369). Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-12189-0\\_31](https://doi.org/10.1007/978-3-642-12189-0_31)

Xie, K. L., Zhang, Z., & Zhang, Z. (2014). The business value of online consumer reviews and management response to hotel performance. *International Journal of Hospitality Management*, 43, 1–12. <https://doi.org/10.1016/J.IJHM.2014.07.007>

Xu, X., Wang, X., Li, Y., & Haghighi, M. (2017). Business intelligence in online customer textual reviews: Understanding consumer perceptions and influential factors. *International Journal of Information Management*, 37(6), 673–683. <https://doi.org/10.1016/J.IJINFOMGT.2017.06.004>

# Anexos

## Anexo A. Acta de Constitución del Proyecto

### Acta de Constitución del Proyecto

---

Nombre del Proyecto:

Departamento:

Fecha:

Director del Proyecto:

---

#### Detalles del Proyecto

##### Visión General del Proyecto

[Reemplazar este texto con una descripción de lo que está involucrado en la ejecución del proyecto. Usar elementos de la lista, si procede.]

- "[Tarea / Proceso / Cambio]"
- "[Tarea / Proceso / Cambio]"
- "[Tarea / Proceso / Cambio]"

##### Problema u Oportunidad

[Reemplazar este texto con una descripción del contexto y los antecedentes del proyecto y por qué es necesario.]

- [Problema/Oportunidad]
- [Problema/Oportunidad]
- [Problema/Oportunidad]

##### Objetivos del Proyecto

[Reemplazar este texto con una descripción de las mejoras que se pueden obtener a través de la implementación de este proyecto y como el entorno empresarial se verá mejorado.]

- [Problema/Oportunidad]
- [Problema/Oportunidad]

## Objetivos Principales del Proyecto

### Objetivos Principales del Proyecto

[Reemplazar este texto con una lista de los elementos clave del proyecto que además de definir y apoyar el proyecto cumplan con el alcance del proyecto.]

- [[Objetivo]
- [[Objetivo]

## Beneficios del Proyecto

### Beneficios del Proyecto

[Describa los beneficios de implementar este proyecto.]

- [Beneficio]
- [Beneficio]

## Entregables Principales del Proyecto

### Hito 1

- [Descripción del Primer Entregable]
- [Descripción del Segundo Entregable]
- [Descripción del Tercer Entregable]

### Hito 2

## Anexos

- [Descripción del Primer Entregable]
- [Descripción del Segundo Entregable]
- [Descripción del Tercer Entregable]

**Hito 3**

- [Descripción del Primer Entregable]
- [Descripción del Segundo Entregable]
- [Descripción del Tercer Entregable]

**No Entregables o Fuera del Alcance de este Proyecto****Fuera del Alcance del Proyecto**

- Especifique lo que esta fuera del alcance del proyecto y por qué.

**Interdependencias del Proyecto****Interdependencias del Proyecto**

[Sustituir este texto con cualquier otro proyecto en proceso que tenga una relación con este proyecto. Incluya una lista de los impactos que pueden tener esos proyectos en el desarrollo de este proyecto.]

- [Impacto]
- [Impacto]

**Condiciones del Proyecto****Presunciones del Proyecto**



[Reemplazar este texto con un resumen de todas las presunciones conocidas que se aplican a este proyecto.]

- [Presunción]
- [Presunción]

### Problemas del Proyecto

[Reemplazar este texto con un resumen de todos los problemas conocidos que se aplican a este proyecto.]

- [Problema]
- [Problema]

### Riesgos del Proyecto

[Reemplazar este texto con un resumen de todos los riesgos conocidos que se aplican a este proyecto.]

- [Riesgo]
- [Riesgo]

### Limitaciones del proyecto

[Reemplazar este texto con un resumen de todas las limitaciones conocidas que se aplican a este proyecto.]

- [Limitacion]
- [Limitacion]

## Factores Críticos para el éxito del proyecto (Indicadores claves de desempeño)

### Factores Críticos para el éxito del Proyecto

[Reemplazar este texto con un resumen de todos los indicadores críticos de éxito que se aplican a este proyecto.]

- [Indicador]
- [Indicador]

## Registro de Interesados en el Proyecto

Interesados del Proyecto
<p>[Reemplazar este texto con un listado de posibles interesados del proyecto]</p> <ul style="list-style-type: none"> <li>• [Interesado]</li> <li>• [Interesado]</li> <li>• [Interesado]</li> </ul>

### Estimación de la Duración del Proyecto

Hitos del Proyecto	Fecha Estimada	Nivel de Certidumbre
Fecha estimada del Comienzo del Proyecto	[mm/dd/aa]	[Alto/Medio/Bajo]
Primer Hito	[mm/dd/aa]	[Alto/Medio/Bajo]
Segundo Hito	[mm/dd/aa]	[Alto/Medio/Bajo]
Tercer Hito	[mm/dd/aa]	[Alto/Medio/Bajo]
Fecha estimada de terminación del Proyecto	[mm/dd/aa]	[Alto/Medio/Bajo]

### Estimación del Presupuesto del Proyecto

Descripción	Duración Estimada	Costo Estimado
Recursos Humanos Internos	[duración]	[costo]
Recursos Humanos Externo	[duración]	[costo]
Equipo (Servidor)	[duración]	[costo]
Herramientas	[duración]	[costo]
Otros	[duración]	[costo]
<b>Total de Costo Estimado</b>		

---

## Aprobaciones

Preparado Por.....

Gerente del Proyecto

Aprobado Por

Patrocinador del proyecto

Cliente Patrocinador (si aplica)